



AA

DEMANDE INTERNATIONALE PUBLIEE EN VERTU DU TRAITE DE COOPERATION EN MATIERE DE BREVETS (PCT)

(51) Classification internationale des brevets ⁶ :		A1	(11) Numéro de publication internationale: WO 97/28186
C07K 14/47, C12N 15/12, C12Q 1/68, A61K 39/395, G01N 33/68			(43) Date de publication internationale: 7 août 1997 (07.08.97)

(21) Numéro de la demande internationale: PCT/FR97/00214	(81) Etats désignés: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, brevet ARIPO (KE, LS, MW, SD, SZ, UG), brevet eurasien (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), brevet européen (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), brevet OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).
(22) Date de dépôt international: 3 février 1997 (03.02.97)	
(30) Données relatives à la priorité: 96/01309 2 février 1996 (02.02.96) FR	
(71) Déposant (<i>pour tous les Etats désignés sauf US</i>): SANOFI [FR/FR]; 32-34, rue Marbeuf, F-75008 Paris (FR).	
(72) Inventeurs; et	Publiée
(75) Inventeurs/Déposants (<i>US seulement</i>): CAPUT, Daniel [FR/FR]; La Bousquière, F-31290 Avignonet-Lauragais (FR). FERRARA, Pascual [AR/FR]; Libouille Saint-Assiscle, F-31290 Avignonet-Lauragais (FR). KAGHAD, Ahmed, Mourad [FR/FR]; 5, rue de la Poste, F-31450 Montgiscard (FR).	<i>Avec rapport de recherche internationale. Avant l'expiration du délai prévu pour la modification des revendications, sera republiée si de telles modifications sont reçues.</i>
(74) Mandataire: LE GUEN, Gérard; Cabinet Lavoix, 2, place d'Estienne-d'Orves, F-75441 Paris Cédex 09 (FR).	

(54) Title: PURIFIED SR-p70 PROTEIN

(54) Titre: PROTEINE PURIFIEE SR-p70

(57) Abstract

Novel nucleic acid sequences from the tumour-suppressor gene family related to the gene of protein p53, and the corresponding protein sequences, are disclosed.

(57) Abrégé

Cette invention a pour objet de nouvelles séquences d'acides nucléiques de la famille des gènes suppresseurs de tumeurs apparentée avec le gène de la protéine p53, et les séquences protéiques correspondantes.

UNIQUEMENT à TITRE D'INFORMATION

Codes utilisés pour identifier les Etats parties au PCT, sur les pages de couverture des brochures publiant des demandes internationales en vertu du PCT.

AT	Arménie	GB	Royaume-Uni	MW	Malawi
AT	Autriche	GE	Géorgie	MX	Mexique
AU	Australie	GN	Guinée	NE	Niger
BB	Barbade	GR	Grèce	NL	Pays-Bas
BE	Belgique	HU	Hongrie	NO	Norvège
BF	Burkina Faso	IE	Irlande	NZ	Nouvelle-Zélande
BG	Bulgarie	IT	Italie	PL	Pologne
BJ	Bénin	JP	Japon	PT	Portugal
BR	Brésil	KE	Kenya	RO	Roumanie
BY	Bélarus	KG	Kirghizistan	RU	Fédération de Russie
CA	Canada	KP	République populaire démocratique de Corée	SD	Soudan
CF	République centrafricaine	KR	République de Corée	SE	Suède
CG	Congo	KZ	Kazakhstan	SG	Singapour
CH	Suisse	LJ	Liechtenstein	SI	Slovénie
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovaquie
CM	Cameroun	LR	Libéria	SN	Sénégal
CN	Chine	LT	Lithuanie	SZ	Swaziland
CS	Tchécoslovaquie	LU	Luxembourg	TD	Tchad
CZ	République tchèque	LV	Lettonie	TG	Togo
DE	Allemagne	MC	Monaco	TJ	Tadjikistan
DK	Danemark	MD	République de Moldova	TT	Trinité-et-Tobago
EE	Estonie	MG	Madagascar	UA	Ukraine
ES	Espagne	ML	Mali	UG	Ouganda
FI	Finlande	MN	Mongolie	US	Etats-Unis d'Amérique
FR	France	MR	Mauritanie	UZ	Ouzbékistan
GA	Gabon			VN	Viet Nam

1

"Protéine purifiée SR-p70".

L'invention concerne de nouvelles séquences d'acides nucléiques de la famille des gènes suppresseurs de tumeurs apparentée avec le gène de la protéine p53, et les séquences protéiques correspondantes.

5 L'invention concerne également les applications prophylactiques, thérapeutiques et diagnostiques de celles-ci, notamment dans le domaine des pathologies liées aux phénomènes d'apoptose ou de transformation cellulaire.

10 Les gènes suppresseurs de tumeurs jouent un rôle clef dans la protection contre les phénomènes de cancérisation, et toute modification susceptible d'entrainer la perte de l'un de ces gènes, son inactivation ou son dysfonctionnement, peut avoir un caractère oncogène, créant ainsi des conditions favorables au développement d'un cancer.

15 Les auteurs de la présente invention ont identifié les produits de transcription d'un nouveau gène ainsi que les protéines correspondantes. Ce gène SR-p70 est apparenté au gène suppresseur de tumeur p53, dont l'activité anti-tumorale est liée à son activité de facteur de transcription et plus spécifiquement aux contrôles exercés sur l'activité des gènes Bax et Bcl-2, instrumentaux dans les mécanismes de mort cellulaire.

20 La présente invention est donc relative à des protéines purifiées SR-p70, ou des fragments biologiquement actifs de celles-ci.

25 L'invention concerne également des séquences d'acides nucléiques isolées codant pour lesdites protéines ou leurs fragments biologiquement actifs et des oligonucléotides spécifiques obtenues à partir de ces séquences.

Elle vise en outre les vecteurs de clonage et/ou d'expression contenant au moins l'une des séquences nucléotidiques définies ci-dessus, et les cellules hôtes transfectées par ces vecteurs de clonage et/ou d'expression dans des conditions permettant la réPLICATION et/ou l'expression de l'une desdites séquences nucléotidiques.

30 Les méthodes de production de protéines recombinantes SR-p70 ou de leurs fragments biologiquement actifs par les cellules hôtes transfectées font également partie de l'invention.

35 L'invention comprend également des anticorps ou des dérivés d'anticorps spécifiques des protéines définies ci-dessus.

Elle vise en outre des méthodes de détection des cancers, soit par la mesure de l'accumulation des protéines SR-p70 dans les tumeurs selon des techniques d'immuno-histochimie, soit par la mise en évidence dans le sérum de patients d'auto-anticorps dirigés contre ces protéines.

L'invention concerne également tout inhibiteur ou activateur de l'activité du SR-p70 par exemple d'interaction protéine-protéine faisant intervenir le SR-p70.

Elle concerne aussi des séquences oligonucléotidiques antisens, spécifiques des séquences d'acides nucléiques ci-dessus, pouvant moduler *in vivo* l'expression du gène SR-p70.

5 L'invention comprend enfin une méthode de thérapie génique dans laquelle des vecteurs tels que par exemple des vecteurs viraux inactivés capables de transférer des séquences codantes pour une protéine selon l'invention sont injectés à des cellules déficientes pour cette protéine, à des fins de régulation des phénomènes 10 d'apoptose ou de réversion de la transformation.

La présente invention a pour objet un polypeptide purifié comprenant une séquence d'acides aminés choisie parmi :

- a) la séquence SEQ ID n° 2 ;
- b) la séquence SEQ ID n° 4 ;
- c) la séquence SEQ ID n° 6 ;
- d) la séquence SEQ ID n° 8 ;
- e) la séquence SEQ ID n° 10 ;
- f) la séquence SEQ ID n° 13 ;
- g) la séquence SEQ ID n° 15 ;
- h) la séquence SEQ ID n° 17 ;
- i) la séquence SEQ ID n° 19 ;
- j) toute séquence biologiquement active dérivée de SEQ ID n° 2, SEQ ID n° 4, SEQ ID n° 6, SEQ ID n° 8, SEQ ID n° 10, SEQ ID n° 13, SEQ ID n° 15, SEQ ID n° 17 ou SEQ ID n° 19.

25 Dans la description de l'invention, on utilise les définitions suivantes :

- protéine SR-p70 : un polypeptide comprenant une séquence d'acides aminés choisie parmi les séquences SEQ ID n° 2, SEQ ID n° 4, SEQ ID n° 6, SEQ ID n° 8, SEQ ID n° 10, SEQ ID n° 13, SEQ ID n° 15, SEQ ID n° 17 ou SEQ ID n° 19, ou tout fragment ou dérivé de celui-ci biologiquement actif.

30 - dérivé : tout polypeptide variant du polypeptide de séquence SEQ ID n° 2, SEQ ID n° 4, SEQ ID n° 6, SEQ ID n° 8, SEQ ID n° 10, SEQ ID n° 13, SEQ ID n° 15, SEQ ID n° 17 ou SEQ ID n° 19 ou toute molécule résultant d'une modification de nature génétique et/ou chimique de la séquence SEQ ID n° 2, SEQ ID n° 4, SEQ ID n° 6, SEQ ID n° 8, SEQ ID n° 10, SEQ ID n° 13, SEQ ID n° 15, SEQ ID n° 17 ou SEQ ID n° 19 c'est-à-dire obtenu par mutation, délétion, addition, substitution et/ou modification chimique d'un seul ou d'un nombre limité d'acides aminés, ainsi qu'une toute séquence

isoforme, c'est-à-dire une séquence identique à la séquence SEQ ID n° 2, SEQ ID n° 4, SEQ ID n° 6, SEQ ID n° 8, SEQ ID n° 10, SEQ ID n° 13, SEQ ID n° 15, SEQ ID n° 17 ou SEQ ID n° 19 à l'un de ses fragments ou séquences modifiées, contenant un ou plusieurs acides aminés sous la forme d'enantiomère D, lesdites séquences variantes, modifiées ou isoformes ayant conservé au moins l'une des propriétés les rendant biologiquement actives.

- biologiquement actif : capable de se lier à l'ADN et/ou d'exercer une activité de facteur de transcription et/ou de participer au contrôle du cycle cellulaire, de la différenciation et de l'apoptose et/ou capable d'être reconnu par les anticorps spécifiques du polypeptide de séquence SEQ ID n° 2, SEQ ID n° 4, SEQ ID n° 6, SEQ ID n° 8, SEQ ID n° 10, SEQ ID n° 13, SEQ ID n° 15, SEQ ID n° 17 ou SEQ ID n° 19 et/ou capable d'induire des anticorps qui reconnaissent ce polypeptide.

La fabrication de dérivés peut avoir différents objectifs, dont en particulier celui d'augmenter l'affinité du polypeptide pour l'ADN ou son activité de facteur de transcription, celui d'améliorer ses taux de production, d'augmenter sa résistance à des protéases, de modifier ses activités biologiques ou de lui conférer de nouvelles propriétés pharmaceutiques et/ou biologiques.

Parmi les polypeptides de l'invention, on préfère le polypeptide d'origine humaine, comprenant la séquence SEQ ID n° 6, SEQ ID n° 13, SEQ ID n° 15, SEQ ID n° 17 ou SEQ ID n° 19. Le polypeptide de 636 acides aminés correspondant à la séquence SEQ ID n° 6 est identique à plus de 97 % au polypeptide de séquence SEQ ID n° 2.

Le polypeptide de séquence SEQ ID n° 2 et celui de séquence SEQ ID n° 4 sont deux produits d'expression d'un même gène, de même pour les séquences SEQ ID n° 8 et SEQ ID n° 10 et pour les séquences SEQ ID n° 6, SEQ ID n° 13, SEQ ID n° 15, SEQ ID n° 17 et SEQ ID n° 19.

Comme il sera expliqué dans les exemples, le polypeptide de séquence SEQ ID n° 4 correspond à une terminaison prématurée du peptide de séquence SEQ ID n° 2, liée à un épissage alternatif du transcript codant pour le polypeptide de SEQ ID n° 2 le plus long (ARN messager) du gène correspondant. De même chez l'humain, les polypeptides correspondant aux séquences SEQ ID n° 6, SEQ ID n° 13, SEQ ID n° 15, SEQ ID n° 17 et SEQ ID n° 19 divergent dans leur composition au niveau des parties, N- et/ou -C terminales et ce consécutif à des épissages alternatifs d'un même transcript primaire. La séquence peptidique N-terminale de la séquence SEQ ID n° 10 est délétée, ce lié à un épissage alternatif de son transcript codant.

Avantageusement, l'invention vise un polypeptide correspondant au domaine de fixation sur l'ADN de l'un des polypeptides précédents.

Ce domaine correspond à la séquence comprise entre le résidu 110 et le résidu 310 pour les séquences SEQ ID n° 2 ou 6, et entre le résidu 60 et le résidu 260 pour la séquence SEQ ID n° 8.

5 La présente invention a également pour objet des séquences d'acides nucléiques codant pour une protéine SR-p70 ou des fragments ou dérivés de celle-ci biologiquement actifs.

Plus préférentiellement, l'invention a pour objet une séquence d'acides nucléiques isolée choisie parmi :

- 10 a) la séquence SEQ ID n° 1 ;
- b) la séquence SEQ ID n° 3 ;
- c) la séquence SEQ ID n° 5 ;
- d) la séquence SEQ ID n° 7 ;
- e) la séquence SEQ ID n° 9 ;
- f) la séquence SEQ ID n° 11 ;
- 15 g) la séquence SEQ ID n° 12 ;
- h) la séquence SEQ ID n° 14 ;
- i) la séquence SEQ ID n° 16 ;
- j) la séquence SEQ ID n° 18 ;
- 20 k) les séquences d'acides nucléiques capables de s'hybrider spécifiquement à la séquence SEQ ID n° 1, SEQ ID n° 3, SEQ ID n° 5, SEQ ID n° 7, SEQ ID n° 9, SEQ ID n° 11, SEQ ID n° 12, SEQ ID n° 14 ou SEQ ID n° 16 ou SEQ ID n° 18 ou à leurs séquences complémentaires, ou de s'hybrider spécifiquement à leurs séquences proximales ;
- 25 l) les séquences dérivées des séquences a), b), c), d), e), f), g), h), i), j) ou k) du fait de la dégénérescence du code génétique.

Selon un mode de réalisation préféré, l'invention a pour objet les séquences nucléotidiqes SEQ ID n° 5, SEQ ID n° 12, SEQ ID n° 14, SEQ ID n° 16 et SEQ ID n° 18 correspondant respectivement aux ADNc des protéines humaines des séquences SEQ ID n° 6, SEQ ID n° 13, SEQ ID n° 15, SEQ ID n° 17 et SEQ ID n° 19.

30 Les différentes séquences nucléotidiqes de l'invention peuvent être d'origine artificielle ou non. Il peut s'agir de séquences d'ADN ou d'ARN, obtenues par criblage de banques de séquences au moyen de sondes élaborées sur la base des séquences SEQ ID n° 1, 3, 5, 7, 9, 11, 12, 14, 16 ou 18. De telles banques peuvent être préparées par des techniques classiques de biologie moléculaire, connues de l'homme de l'art.

Les séquences nucléotidiques selon l'invention peuvent également être préparées par synthèse chimique, ou encore par des méthodes mixtes incluant la modification chimique ou enzymatique de séquences obtenues par criblage de banques.

5 Ces séquences nucléotidiques permettent la réalisation de sondes nucléotidiques, capables de s'hybrider fortement et spécifiquement avec une séquence d'acides nucléiques, d'un ADN génomique ou d'un ARN messager, codant pour un polypeptide selon l'invention ou un fragment biologiquement actif de celui-ci. De telles sondes font également partie de l'invention. Elles peuvent être utilisées comme outil de diagnostic *in vitro* pour la détection, par des expériences d'hybridation, de transcripts spécifiques des polypeptides de l'invention dans des échantillons biologiques ou pour la mise en évidence de synthèses aberrantes ou d'anomalies génétiques telles que la perte d'hétérozygotie ou le réarrangement génétique, résultant d'un polymorphisme, de mutations ou d'un épissage différent.

10 Les sondes de l'invention comportent au minimum 10 nucléotides, et au maximum comportent la totalité de la séquence du gène SR-p70 ou de son ADNc contenu par exemple dans un cosmide.

15 Parmi les sondes les plus courtes, c'est-à-dire d'environ 10 à 20 nucléotides, les conditions d'hybridation appropriées correspondent aux conditions stringentes usuellement utilisées par l'homme de métier.

20 La température utilisée est de préférence comprise entre T_m -5° C à T_m -30° C, de préférence encore entre T_m -5° C et T_m -10° C, T_m étant la température de fusion, température à laquelle 50 % des brins d'ADN appariés se séparent.

25 L'hybridation est de préférence menée dans des solutions à force ionique élevée, telles que notamment des solutions 6 x SSC.

De manière avantageuse, les conditions d'hybridation utilisées sont les suivantes :
- température : 42° C,
- tampon d'hybridation : 6 x SSC, 5 x Denhart's, 0,1 % SDS,
telles que décrites dans l'exemple III.

30 Avantageusement, ces sondes sont représentées par les oligonucléotides suivants ou leurs complémentaires :

SEQ ID n° 20 : GCG AGC TGC CCT CGG AG
SEQ ID n° 21 : GGT TCT GCA GGT GAC TCA G
SEQ ID n° 22 : GCC ATG CCT GTC TAC AAG
SEQ ID n° 23 : ACC AGC TGG TTG ACG GAG
SEQ ID n° 24 : GTC AAC CAG CTG GTG GGC CAG
SEQ ID n° 25 : GTG GAT CTC GGC CTC C

SEQ ID n° 26 : AGG CCG GCG TGG GGA AG
SEQ ID n° 27 : CTT GGC GAT CTG GCA GTA G
SEQ ID n° 28 : GCG GCC ACG ACC GTG AC
SEQ ID n° 29 : GGC AGC TTG GGT CTC TGG
5 SEQ ID n° 30 : CTG TAC GTC GGT GAC CCC
SEQ ID n° 31 : TCA GTG GAT CTC GGC CTC
SEQ ID n° 32 : AGG GGA CGC AGC GAA ACC
SEQ ID n° 33 : CCA TCA GCT CCA GGC TCT C
SEQ ID n° 34 : CCA GGA CAG GCG CAG ATG
10 SEQ ID n° 35 : GAT GAG GTG GCT GGC TGG A
SEQ ID n° 36 : TGG TCA GGT TCT GCA GGT G
SEQ ID n° 37 : CAC CTA CTC CAG GGA TGC
SEQ ID n° 38 : AGG AAA ATA GAA GCG TCA GTC
SEQ ID n° 39 : CAG GCC CAC TTG CCT GCC
15 SEQ ID n° 40 : CTG TCC CCA AGC TGA TGA G

Préférentiellement, les sondes de l'invention sont marquées, préalablement à leur utilisation. Pour cela, plusieurs techniques sont à la portée de l'homme du métier (marquage fluorescent, radioactif, chimioluminescent, enzymatique, etc).
20 Les méthodes de diagnostic *in vitro* dans lesquelles ces sondes nucléotidiques sont mises en œuvre, sont incluses dans l'objet de la présente invention.
Ces méthodes concernent par exemple la détection de synthèses anormales (ex. accumulation de produits de transcription) ou d'anomalies génétiques, telles que la perte d'hétérozygotie et le réarrangement génétique, et les mutations ponctuelles au niveau des séquences nucléotidiques d'acides nucléiques codant pour une protéine SR-p70, selon la définition donnée précédemment.
25 Les séquences nucléotidiques de l'invention sont également utiles pour la fabrication et l'utilisation d'amorces oligonucléotidiques pour des réactions de séquençage ou d'amplification spécifique selon la technique dite de PCR ou toute variante de celle-ci (Ligase Chain Reaction (LCR), ...).
Des paires d'amorces préférées sont constituées par des amorces choisies sur les séquences nucléotidiques : SEQ ID n° 1 : séquence de singe de 2 874 nucléotides et SEQ ID n° 5 : ADNc SR-p70a humain, notamment en amont du codon ATG d'initiation et en aval du codon TGA d'arrêt de traduction.
30 35 Avantageusement, ces amorces sont représentées par les couples suivants:

- coupl n°1 :

amorce sens : GCG AGC TGC CCT CGG AG (SEQ ID n° 20)

amorce antisens : GGT TCT GCA GGT GAC TCA G (SEQ ID n° 21)

5

- couple n°2 :

amorce sens : GCC ATG CCT GTC TAC AAG (SEQ ID n° 22)

amorce antisens : ACC AGC TGG TTG ACG GAG (SEQ ID n° 23)

10

- couple n° 3 :

amorce sens : GTC AAC CAG CTG GTG GGC CAG (SEQ ID n° 24)

amorce antisens : GTG GAT CTC GGC CTC C (SEQ ID n° 25)

15

- couple n° 4 :

amorce sens : AGG CCG GCG TGG GGA AG (SEQ ID n° 26)

amorce antisens : CTT GGC GAT CTG GCA GTA G (SEQ ID n° 27)

20

- couple n° 5 :

amorce sens : GCG GCC ACG ACC GTG A (SEQ ID n° 28)

amorce antisens : GGC AGC TTG GGT CTC TGG (SEQ ID n° 29)

25

- couple n° 6 :

amorce sens : CTG TAC GTC GGT GAC CCC (SEQ ID n° 30)

amorce antisens : TCA GTG GAT CTC GGC CTC (SEQ ID n° 31)

30

- couple n° 7 :

amorce sens : AGG GGA CGC AGC GAA ACC (SEQ ID n° 32)

amorce antisens : GGC AGC TTG GGT CTC TGG (SEQ ID n° 29)

35

- couple n° 8 :

amorce sens : CCCCCCCCCCCCN (où N est égal à G, A ou T)

amorce antisens : CCA TCA GCT CCA GGC TCT C (SEQ ID n° 33)

- couple n° 9 :

amorce sens : CCCCCCCCCCCCN (où N est égal à G, A ou T)

amorce antisens : CCA GGA CAG GCG CAG ATG (SEQ ID n° 34)

- couple n° 10 :

amorce sens : CCCCCCCCCCCCN (où N est égal à G, A ou T)

amorce antisens : CTT GGC GAT CTG GCA GTA G (SEQ ID n° 27)

5

- couple n° 11 :

amorce sens : CAC CTA CTC CAG GGA TGC (SEQ ID n° 37)

amorce antisens : AGG AAA ATA GAA GCG TCA GTC (SEQ ID n° 38)

10

- couple n° 12 :

amorce sens : CAG GCC CAC TTG CCT GCC (SEQ ID n° 39)

amorce antisens : CTG TCC CCA AGC TGA TGA G (SEQ ID n° 40)

Ces amorces correspondent aux séquences allant respectivement :

- du nucléotide n° 124 au nucléotide n° 140 sur SEQ ID n° 1 et du nucléotide n° 1 au nucléotide n° 17 sur SEQ ID n° 5 pour SEQ ID N° 20
- du nucléotide n° 2280 au nucléotide n° 2262 sur SEQ ID n° 1 et du nucléotide n° 2156 au nucléotide 2138 sur SEQ ID n° 5 pour SEQ ID N° 21
- du nucléotide n° 684 au nucléotide n° 701 sur SEQ ID n° 1 pour SEQ ID N° 22
- du nucléotide n° 1447 au nucléotide n° 1430 sur SEQ ID n° 1 et du nucléotide 1324 au nucléotide 1307 sur SEQ ID n° 5 pour SEQ ID N° 23
- du nucléotide 1434 au nucléotide 1454 sur SEQ ID n° 1 et du nucléotide 1311 au nucléotide 1331 sur SEQ ID n° 5 pour SEQ ID n° 24
- du nucléotide 2066 au nucléotide 2051 sur SEQ ID n° 1 et du nucléotide 1940 au nucléotide 1925 sur SEQ ID n° 5 pour SEQ ID n° 25.
- du nucléotide 18 au nucléotide 32 sur SEQ ID n° 5 pour SEQ ID n° 26
- du nucléotide 503 au nucléotide 485 sur SEQ ID n° 5 pour SEQ ID n° 27
- du nucléotide 160 au nucléotide 176 sur SEQ ID n° 11 pour SEQ ID n° 28
- du nucléotide 1993 au nucléotide 1976 sur SEQ ID n° 5 pour SEQ ID n° 29
- du nucléotide 263 au nucléotide 280 sur SEQ ID n° 11 pour SEQ ID n° 30
- du nucléotide 1943 au nucléotide 1926 sur SEQ ID n° 5 pour SEQ ID n° 31
- du nucléotide 128 au nucléotide 145 sur la séquence nucléotidique représentée à la figure 22 pour SEQ ID n° 32
- du nucléotide 1167 au nucléotide 1149 sur SEQ ID n° 5 pour SEQ ID n° 33
- du nucléotide 928 au nucléotide 911 sur SEQ ID n° 5 pour SEQ ID n° 34
- du nucléotide 677 au nucléotide 659 sur SEQ ID n° 5 pour SEQ ID n° 35
- du nucléotide 1605 au nucléotide 1587 sur SEQ ID n° 5 pour SEQ ID n° 36

- du nucléotide 1 au nucléotide 18 sur la séquence nucléotidique représentée à la figure 13 pour SEQ ID n° 37
 - du nucléotide 833 au nucléotide 813 sur la séquence nucléotidique représentée à la figure 13 pour SEQ ID n° 38
 - 5 - du nucléotide 25 au nucléotide 42 sur la séquence nucléotidique représentée à la figure 13 pour SEQ ID n° 39
 - du nucléotide 506 au nucléotide 488 sur la séquence nucléotidique représentée à la figure 13 pour SEQ ID n° 40
- 10 Les séquences nucléotidiques selon l'invention peuvent avoir par ailleurs des utilisations en thérapie génique, notamment pour le contrôle des phénomènes d'apoptose et de réversion de la transformation.
- 15 Les séquences nucléotidiques selon l'invention peuvent par ailleurs être utilisées pour la production de protéines recombinantes SR-p70, selon la définition qui a été donnée à ce terme.
- 20 Ces protéines peuvent être produites à partir des séquences nucléotidiques définies ci-dessus, selon des techniques de production de produits recombinants connues de l'homme du métier. Dans ce cas, la séquence nucléotidique utilisée est placée sous le contrôle de signaux permettant son expression dans un hôte cellulaire.
- 25 Un système efficace de production d'une protéine recombinante nécessite de disposer d'un vecteur, par exemple d'origine plasmidique ou virale, et d'une cellule hôte compatible.
- L'hôte cellulaire peut être choisi parmi des systèmes procaryotes, comme les bactéries, ou eucaryotes, comme par exemple les levures, cellules d'insectes, CHO (cellules d'ovaires de hamster chinois) ou tout autre système avantageusement disponible. Un hôte cellulaire préféré pour l'expression des protéines de l'invention est constitué par la bactérie *E. coli*, notamment la souche MC 1061 (Clontec).
- 30 Le vecteur doit comporter un promoteur, des signaux d'initiation et de terminaison de la traduction, ainsi que les régions appropriées de régulation de la transcription. Il doit pouvoir être maintenu de façon stable dans la cellule et peut éventuellement posséder des signaux particuliers spécifiant la sécrétion de la protéine traduite.
- Ces différents signaux de contrôle sont choisis en fonction de l'hôte cellulaire utilisé. A cet effet, les séquences nucléotidiques selon l'invention peuvent être insérées dans des vecteurs à réPLICATION autonome au sein de l'hôte choisi, ou des vecteurs intégratifs de l'hôte choisi. De tels vecteurs seront préparés selon les méthodes couramment utilisées par l'homme du métier, et les clones en résultant peuvent être

introduits dans un hôte approprié par des méthodes standard, telles que par exemple l'électroporation.

Les vecteurs de clonage et/ou d'expression contenant au moins l'une des séquences nucléotidiques définies ci-dessus font également partie de la présente invention.

5 Un vecteur de clonage et d'expression préféré est le plasmide pSE1 qui comporte à la fois les éléments nécessaires pour son utilisation comme vecteur de clonage dans *E.coli* (origine de réplication dans *E. coli* et gène de résistance à l'ampicilline, provenant du plasmide pTZ 18R), et comme vecteur d'expression dans les cellules animales (promoteur, intron, site de polyadenylation, origine de réplication du virus SV40), ainsi que les éléments permettant sa copie en simple brin dans un but de séquençage (origine de réplication du phage f1).

Les caractéristiques de ce plasmide sont décrites dans la demande EP 0 506 574.

10 Sa construction, ainsi que l'intégration des ADNc provenant des séquences d'acides nucléiques de l'invention sont par ailleurs décrites dans les exemples ci-après.

15 Selon un mode de réalisation préféré, les protéines de l'invention sont sous forme de protéines de fusion, notamment sous forme de protéine fusionnée avec la glutathione S-transférase (GST). Un vecteur d'expression désigné dans ce cas est représenté par le vecteur plasmidique pGEX-4T-3 (Pharmacia ref-27.4583).

20 L'invention vise en outre les cellules hôtes transfectées par ces vecteurs précédents. Ces cellules peuvent être obtenues par l'introduction dans des cellules hôtes d'une séquence nucléotidique insérée dans un vecteur tel que défini ci-dessus, puis la mise en culture desdites cellules dans des conditions permettant la réplication et/ou l'expression de la séquence nucléotidique transfectée.

25 Ces cellules sont utilisables dans une méthode de production d'un polypeptide recombinant de séquence SEQ ID n° 2, SEQ ID n° 4, SEQ ID n° 6, SEQ ID n° 8, SEQ ID n° 10, SEQ ID n° 12, SEQ ID n° 14, SEQ ID n° 16 ou SEQ ID n° 18 ou tout fragment ou dérivé biologiquement actif de celui-ci.

30 La méthode de production d'un polypeptide de l'invention sous forme recombinante est elle-même comprise dans la présente invention, et se caractérise en ce que l'on cultive les cellules transfectées dans des conditions permettant l'expression d'un polypeptide recombinant de séquence SEQ ID n° 2, SEQ ID n° 4, SEQ ID n° 6, SEQ ID n° 8, SEQ ID n° 10, SEQ ID n° 12, SEQ ID n° 14, SEQ ID n° 16 ou SEQ ID n° 18 ou de tout fragment ou dérivé biologiquement actif de celui-ci, et que l'on récupère ledit polypeptide recombinant.

35 Les procédés de purification utilisés sont connus de l'homme du métier. Le polypeptide recombinant peut être purifié à partir de lysats et extraits cellulaires, du

- 5 surnageant du milieu de culture, par des méthodes utilisées individuellement ou en combinaison, telles que le fractionnement, les méthodes de chromatographie, les techniques d'immunoaffinité à l'aide d'anticorps mono ou polyclonaux spécifiques, etc. Une variante préférée consiste à produire un polypeptide recombinant fusionné à une protéine "porteuse" (protéine chimère). L'avantage de ce système est qu'il permet une stabilisation et une diminution de la protéolyse du produit recombinant, une augmentation de la solubilité au cours de la renaturation *in vitro* et/ou une simplification de la purification lorsque le partenaire de fusion possède une affinité pour un ligand spécifique.
- 10 10 Avantageusement, les polypeptides de l'invention sont fusionnés avec la glutathion S-transférase en position N-terminale (système "GST" Pharmacia). Le produit de fusion est dans ce cas détecté et quantifié grâce à l'activité enzymatique de la GST. Le réactif colorimétrique utilisé est un accepteur de glutathion, substrat de la GST. Le produit recombinant est purifié sur un support de chromatographie auquel ont été préalablement couplées des molécules de glutathion.
- 15 15 Les anticorps mono ou polyclonaux capables de reconnaître spécifiquement une protéine SR-p70 selon la définition donnée précédemment font également partie de l'invention. Des anticorps polyclonaux peuvent être obtenus à partir du sérum d'un animal immunisé contre la protéine, produite par exemple par recombinaison génétique suivant la méthode décrite ci-dessus, selon les modes opératoires usuels. Les anticorps monoclonaux peuvent être obtenus selon la méthode classique de culture d'hybridomes décrite par Köhler et Milstein, Nature, 1975, 256, 495-497.
- 20 20 Des anticorps avantageux sont des anticorps dirigés contre la région centrale comprise entre le résidu 110 et le résidu 310 pour les séquences SEQ ID n° 2 ou 6 ou entre le résidu 60 et le résidu 260 pour la séquence SEQ ID n° 8.
- 25 25 Les anticorps selon l'invention sont par exemple des anticorps chimériques, des anticorps humanisés, des fragments Fab et F(ab')2. Ils peuvent également se présenter sous forme d'immunoconjugués ou d'anticorps marqués.
- 30 30 Par ailleurs, outre leur utilisation pour la purification des polypeptides recombinants, les anticorps de l'invention, en particulier les anticorps monoclonaux, peuvent également être utilisés pour la détection de ces polypeptides dans un échantillon biologique.
- 35 35 Ils constituent ainsi un moyen d'analyse immunocytochimique ou immunohistochimique de l'expression de protéines SR-p70 sur des coupes de tissus

spécifiques, par exemple par immunofluorescence, marquage à l'or, immunoconjugués enzymatiques.

Il 5 permettent notamment de mettre en évidence une accumulation anormale de protéines SR-p70 dans certains tissus ou prélèvements biologiques, ce qui les rend utiles pour la détection des cancers ou le suivi de l'évolution ou de la rémission de cancers préexistants.

Plus généralement, les anticorps de l'invention peuvent être avantageusement mis en oeuvre dans toute situation où l'expression d'une protéine SR-p70 doit être observée. L'invention concerne donc également un procédé de diagnostic *in vitro* de pathologies 10 corrélées à une expression ou une accumulation anormale de protéines SR-p70, notamment les phénomènes de cancérisation, à partir d'un prélèvement biologique, caractérisé en ce que l'on met en contact au moins un anticorps de l'invention avec ledit prélèvement biologique, dans des conditions permettant la formation éventuelle de complexes immunologiques spécifiques entre une protéine SR-p70 et le ou lesdits anticorps et en ce que l'on détecte les complexes immunologiques spécifiques éventuellement formés.

L'invention concerne également un kit pour le diagnostic *in vitro* d'une expression ou 20 une accumulation anormale de protéines SR-p70 dans un prélèvement biologique et/ou pour la mesure du taux d'expression de celle-ci dans ledit prélèvement comprenant :

- au moins un anticorps spécifique d'une protéine SR-p70, éventuellement fixé sur un support,
- des moyens de révélation de la formation de complexes antigènes/anticorps spécifiques entre une protéine SR-p70 et ledit anticorps et/ou des moyens de quantification de ces complexes.

L'invention vise également une méthode de diagnostic précoce de la formation des tumeurs, par la mise en évidence dans le sérum d'un individu, d'auto-anticorps dirigés contre une protéine SR-p70.

Une telle méthode de diagnostic précoce est caractérisée en ce que l'on met en contact un échantillon de sérum prélevé chez un individu avec un polypeptide de l'invention, éventuellement fixé sur un support, dans des conditions permettant la formation de complexes immunologiques spécifiques entre ledit polypeptide et les auto-anticorps éventuellement présents dans l'échantillon de sérum, et en ce que l'on détecte les complexes immunologiques spécifiques éventuellement formés.

35 L'invention a également pour objet une méthode de détermination d'une variabilité allélique, d'une mutation, d'une délétion, d'une insertion, d'une perte d'hétérozygotie

ou d'une anomalie génétique du gène SR-p70, pouvant être impliquées dans des pathologies caractérisée en ce qu'elle utilise au moins une séquence nucléotidique décrite ci-dessus. Parmi les méthodes de détermination d'une variabilité allélique, d'une mutation, d'une délétion, d'une insertion, d'une perte d'hétérozygotie ou d'une
5 anomalie génétique du gène SR-p70, on préfère la méthode caractérisée en ce qu'elle comprend au moins une étape d'amplification par PCR de la séquence nucléotidique cible du SR-p70 susceptible de présenter un polymorphisme, une mutation, une délétion ou une insertion à l'aide de couple d'amorces de séquences nucléotidiques définies ci-dessus, une étape au cours de laquelle on procède au traitement des produits amplifiés à l'aide d'enzyme de restriction approprié et une
10 étape au cours de laquelle on procède à la détection ou au dosage d'au moins l'un des produits de la réaction enzymatique.

L'invention comprend également des compositions pharmaceutiques comprenant comme principe actif un polypeptide répondant aux définitions précédentes, préférentiellement sous forme soluble, associé à un véhicule pharmaceutiquement acceptable.
15

De telles compositions offrent une nouvelle approche pour traiter les phénomènes de cancérisation au niveau du contrôle de la multiplication et la différenciation cellulaire. Préférentiellement, ces compositions peuvent être administrées par voie systémique,
20 de préférence par voie intraveineuse, par voie intramusculaire, intradermique ou par voie orale.

Leurs modes d'administration, posologies et formes galéniques optimaux peuvent être déterminés selon les critères généralement pris en compte dans l'établissement d'un traitement thérapeutique adapté à un patient comme par exemple l'âge ou le poids corporel du patient, la gravité de son état général, la tolérance au traitement et les effets secondaires constatés, etc.
25

L'invention comprend enfin une méthode de thérapie génique dans laquelle des séquences nucléotidiques codant pour une protéine SR-p70 sont transférées à des cellules cibles par le biais de vecteurs viraux inactivés.
30

D'autres caractéristiques et avantages de l'invention apparaissent dans la suite de la description avec les exemples et les figures dont les légendes sont représentées ci-après.

LEGENDE DES FIGURES

- Figure 1 : Comparaison nucléique de l'ADNc du SR-p70a de singe (correspondant à SEQ ID n°1) avec la séquence nucléique de l'ADNc de p53 de singe.
- 5
- Figure 2 : Comparaison protéique de SR-p70a de singe avec la protéine p53 de singe (sw : p53-cerae).
- 10
- Figure 3 : Comparaison de la séquence nucléique de l'ADNc de SR-p70a et b de singe (correspondant respectivement à SEQ ID n° 1 et SEQ ID n° 3).
- Figure 4 : Séquence nucléique et séquence protéique déduite de SR-p70a de singe.
- 15
- Figure 5 : Séquence nucléique partielle et séquence protéique déduite complète de SR-p70b de singe.
- 20
- Figure 6 : Séquence nucléique partielle et séquence protéique complète déduite de SR-p70a humain (correspondant à SEQ ID n° 5).
- Figure 7 : Séquence nucléique partielle et séquence protéique déduite complète de SR-p70c de souris (correspondant à SEQ ID n° 7).
- 25
- Figure 8 : Séquence nucléique partielle et séquence protéique déduite partielle de SR-p70a de souris (correspondant à SEQ ID n° 9).
- 30
- Figure 9 : Multialignement des protéines déduites des ADNc SR-p70 de singe (a et b), humain (a) et de souris (a et c).
- Figure 10a : Immunoempreinte de la protéine SR-p70.
- 35
- Figure 10b : Détection de la protéine endogène SR-p70.

Figure 11 : Localisation chromosomique du gène SR-p70 humain. Le signal apparaît sur le chromosome 1, dans la région p36.

5 Figure 12 : Structure génomique du gène SR-p70 et comparaison avec celle du gène p53. Les séquences protéiques humaines du SR-p70a (ligne du haut de l'alignement) et de la p53 (ligne du bas) sont morcelées en peptides en fonction des exons respectifs à partir desquels ils sont codés. Les chiffres au niveau des flèches correspondent à la numérotation des exons correspondants.

10

15 Figure 13 : Séquence génomique humaine du SR-p70 depuis le 3' de l'intron 1 jusqu'au 5' de l'exon 3. Les introns sont encadrés. Aux positions 123 et 133, sont localisées deux positions nucléiques variables (G → A en 123 et C → T en 133). Les sites de restriction de l'enzyme StyI sont soulignés (position 130 dans le cas où il y a présence d'un T au lieu d'un C à la position 133, position 542 et position 610). Les flèches positionnent les amorces nucléiques utilisées dans l'exemple XI.

20 Figure 14 : Comparaison nucléique du 5' des ADNc humains du SR-p70d et du SR-p70a.

25

Figure 15 : Multialignment des séquences nucléiques correspondant au SR-p70 humain a, b, d, e, et f.

30 Figure 16 : Multi-alignement des protéines déduites des ADNc SR-p70 humains (a, b, d, e et f).

35 Figure 17 : Séquence nucléique partielle et séquence protéique déduite partielle du SR-p70a humain. Les deux bases en caractères gras correspondent à deux positions variables (voir figure 6). Cette séquence présente une région 5' non codante plus complète que celle présentée dans la figure 6.

Figure 18 : Analyse des transcrits SR-p70a après amplification par PCR.
piste M : marqueurs de poids moléculaires "1 kb ladder" (GIBCO-BRL)

piste 1 : lignée HT29
piste 3 : lignée SK-N-AS
piste 5 : lignée UMR-32
piste 7 : lignée U-373 MG
piste 9 : lignée SW 480
piste 11 : lignée CHP 212
piste 13 : lignée SK-N-MC

5

10 pistes 2, 4, 6, 8, 10, 12, 14 ; témoins négatifs correspondant aux pistes 1, 3, 5, 7, 9, 11 et 13 respectivement (absence de transcriptase inverse dans la réaction RT-PCR).

15 Analyse par électrophorèse sur gel d'agarose des fragments génomiques amplifiés par PCR (depuis le 3' de l'intron 1 jusqu'au 5' de l'exon 3). La numérotation des pistes correspond à la numérotation de l'échantillonnage témoin. Piste M : marqueurs de poids moléculaires ("1 kb ladder").

B : Analyse identique à celle de la partie A après une digestion par l'enzyme de restriction StyI des mêmes échantillons.

20

Figure 20 : Représentation schématique avec une carte de restriction partielle du plasmide pCDNA3 contenant le SR-p70a humain.

25

30

35

EXEMPLE I

Clonage de l'ADNc du SR-p70 de cellules COS-3.

5 *1. Culture des cellules COS-3*

Les cellules COS-3 (cellules de rein de singe vert d'Afrique transformées par l'antigène T du virus SV 40) sont cultivées dans le milieu DMEM (GIBCO-BRL référence 41 965-047) contenant 2 mM de L-glutamine et supplémenté avec 50 mg/l de gentamycine et de 5 % de sérum de bovin foetal (GIBCO-BRL référence 10231-074) jusqu'à semi-confluence.

10 *2. Préparation de l'ARN messager*

15 a) extraction de l'ARN messager

Les cellules sont récupérées de la façon suivante :

15 - les cellules adhérentes sont lavées deux fois avec du tampon PBS (phosphate buffered saline, référence 04104040-GIBCO-BRL) puis grattées avec un grattoir en caoutchouc et centrifugées.

20 Le culot cellulaire est mis en suspension dans le tampon de lyse de composition suivante : guanidine-thiocyanate 4M ; citrate de sodium 25mM pH 7 ; sarcosyl 0,5 % ; β-mercaptopropanoïde 0,1 M. La suspension est soniquée à l'aide d'un sonicateur Ultra-Turrax n° 231256(Janke et Kundel) à puissance maximale pendant une minute. On ajoute de l'acétate de sodium pH 4 jusqu'à 0,2 M. La solution est extraite avec un volume d'un mélange phénol/chloroforme (v/v ; 5/1). On précipite à -20°C l'ARN contenu dans la phase aqueuse à l'aide d'un volume d'isopropanol. Le culot est resuspendu dans le tampon de lyse. La solution est à nouveau extraite avec un mélange phénol/chloroforme et l'ARN est précipité avec de l'isopropanol. Après lavage du culot avec de l'éthanol 70 % puis 100 %, l'ARN est resuspendu dans de l'eau.

25 b) Purification de la fraction poly A⁺ de l'ARN

30 La purification de la fraction poly A⁺ de l'ARN est réalisée à l'aide du kit Dynabeads oligo (dT)₂₅ de DYNAL (référence 610.05) suivant le protocole préconisé par le fabricant. Le principe est basé sur l'utilisation de billes polystyrène superparamagnétique sur lesquelles est attaché un oligonucléotide poly(dT)₂₅. La fraction poly A⁺ de l'ARN est hybridée sur l'oligo(dT)₂₅ couplé aux billes que l'on piège sur un support magnétique .

3. Constitution de la banque d'ADN complémentaire

a) préparation de l'ADN complémentaire

5 A partir de 0,5 µg des ARN-poly A⁺ de cellules COS-3 obtenus à l'issue de l'étape 2, on prépare l'ADN complémentaire simple-brin marqué au ³²P.dCTP (l'ADN complémentaire obtenu présente une activité spécifique de 3000 dpm/ng) avec l'amorce synthétique de séquence suivante (comprenant un site BamHI) :

5'-<GATCCGGGCC CTTTTTTTTT TTT<3'

10 dans un volume de 30 µl de tampon de composition : Tris HCl 50 mM pH 8,3, MgCl₂ 6 mM, DTT 10 mM, KCl 40 mM, contenant 0,5 mM de chacun des désoxynucléotides triphosphates, 30µCi de dCTP α^{32} P et 30 U de RNasin (promega). Après une heure d'incubation à 37°C, puis 10 minutes à 50°C, puis de nouveau 10 minutes à 37°C, avec 200 unités de l'enzyme transcriptase inverse RNase H⁻ (GIBCO-BRL référence 8064A), on ajoute 4 µl d'EDTA.

b) Hydrolyse alcaline de la matrice ARN

15 On ajoute 6 µl d'une solution de NaOH 2N, puis on incube pendant 5 minutes à 65° C.

c) Purification sur colonne sephacryl S400

20 Afin d'éliminer l'amorce synthétique, on purifie l'ADN complémentaire sur une colonne de 1 ml de sephacryl S400 (Pharmacia), équilibrée dans du tampon TE. Les deux premières fractions radioactives sont regroupées et précipitées avec 1/10 de volume d'une solution d'acétate d'ammonium 10 M et 2,5 volumes d'éthanol, ceci après une extraction, avec un volume de chloroforme.

d) Addition homopolymérique de dG

25 On allonge l'ADN complémentaire en 3' avec une "queue" de dG avec 20 unités de l'enzyme terminale transférase (Pharmacia 27073001). On incube dans 20 µl de tampon de composition : Tris HCl 30 mM pH 7,6 ; chlorure de cobalt 1mM, acide cacodylique 140 mM, DTT 0,1mM, dGTP 1 mM, pendant 15 minutes à 37°C, puis on ajoute 2 µl d'EDTA 0,5 M.

e) On répète à nouveau les étapes b) et c)

30 f) Appariement du vecteur de clonage pSE1 (EP 506 574) et de l'ADN complémentaire en présence de l'adaptateur.

On centrifuge, le culot est dissous dans 33 µl de tampon TE, on ajoute 5 µl (125 ng) de vecteur de clonage pSE1, 1 µl(120 ng) de l'adaptateur de séquence suivante (comprenant un site Apal) :

35 5'AAAAAAAAAAAAAGGGCCCG3'.

10 µl d'un solution de NaCl 200 mM, on incube pendant 5 minutes à 65°C puis on laisse refroidir le mélange réactionnel jusqu'à température ambiante.

g) Ligation

On lie le vecteur de clonage et l'ADNc simple brin dans un volume de 100 µl avec 5 32,5 unités de l'enzyme ADN ligase du phage T4 (Pharmacia référence 270 87002) pendant une nuit à 15°C dans un tampon de composition : Tris HCl 50 mM pH 7,5, MgCl₂ 10 mM, ATP 1 mM.

h) Synthèse du deuxième brin de l'ADNc

On élimine les protéines par extraction au phénol suivie d'une extraction au 10 chloroforme, puis on ajoute 1/10ème de volume d'une solution d'acétate d'ammonium 10 mM, puis 2,5 volumes d'éthanol. On centrifuge, le culot est dissous dans le tampon de composition Tris acétate 33 mM pH 7,9 acétate de potassium 62,5 mM, acétate de magnésium 1 mM et dithiothréitol (DTT) 1 mM, le deuxième brin d'ADN complémentaire est synthétisé dans un volume de 30 µl avec 30 unités de l'enzyme 15 ADN polymérase du phage T4 (Pharmacia, référence 270718) et un mélange de 1 mM des quatre désoxynucléotides triphosphates de dATP, dCTP, dGTP et dTTP, ainsi que deux unités de la protéine du gène 32 du phage T4 (Pharmacia, référence 27-0213)) pendant une heure à 37°C. On extrait au phénol et on retire les traces de 20 phénol par une colonne de polyacrylamide P10 (Biogel P10-200-400 mesh - référence 15011050 - Biorad).

i) Transformation par électroporation

On transforme des cellules *E. Coli* MC 1061 avec l'ADN recombinant obtenu 25 précédemment par électroporation à l'aide de l'appareil Biorad Gene Pulser (Biorad) utilisé à 2,5 kV dans les conditions prescrites par le fabricant, puis on fait pousser les bactéries pendant une heure dans du milieu dit milieu LB (Sambrook *op cité*) de composition : bactotryptone 10 g/l ; extrait de levure 5 g/l ; NaCl 10 g/l.

On détermine le nombre de clones indépendants en étalant une dilution au 1/1000ème de la transformation après la première heure d'incubation sur une boîte 30 de milieu LB additionné de 1,5 % d'agar (p/v) et de 100 µg/ml d'ampicilline, appelé par la suite milieu LB gélosé. Le nombre de clones indépendants est de 1 million.

j) Analyse des ADNc de la banque

Dans le cadre de l'analyse de clones individualisés de la banque par un 35 séquençage nucléique du 5' des ADNc, un clone, dénommé SR-p70a s'est révélé présenter une homologie partielle avec l'ADNc de la protéine déjà connue, la protéine p53 (Genbank X 02469 et X 16384) (Figure 1). Les séquences ont été réalisées avec le kit United States Biochemical (référence 70770) et/ou le kit Applied

Biosystems (références 401434 et/ou 401628) qui utilisent la méthode de Sanger et al., Proc. Natl. Acad. Sci. USA ; 1977, 74, 5463-5467. L'ADN plasmidique est préparé à partir du kit WIZARD mini préparation (Promega référence A7510). Les amorces utilisées sont des oligonucléotides de 16 à 22 mer, complémentaires soit au vecteur pSE1 dans la région immédiatement en 5' de l'ADNc, soit à la séquence de l'ADNc.

Un second ADNc a été isolé à partir de la même banque en criblant de manière similaire à la technique décrite dans l'EXEMPLE III 3) ci-après avec un fragment de l'ADN SR-p70a marqué au ^{32}P avec le kit BRL "Random Primers DNA labelling systems" (référence 18187-013). Les tampons d'hybridation et de lavage sont additionnés de 50 % de formamide. Le dernier lavage est réalisé en 0,1 x SSC/SDS 0,1 % à 60°C. Cette seconde séquence (ADNc SR-p70b) est identique à la première mais présente un fragment interne déleté (Figure 3).

Les deux ADNc SR-p70, d'une longueur de 2874 nucléotides (SR-p70a) et de 2780 nucléotides (SR-p70b) correspondent aux produits d'un seul gène, un épissage alternatif entraînant une déletion de 94 bases entre les nucléotides 1637 et 1732 et une terminaison prématurée de la protéine codée correspondante. Les protéines déduites des deux ADNc présentent respectivement 637 acides aminés et 499 acides aminés (Figures 4 et 5).

20

EXEMPLE II

Obtention de la séquence et clonage de l'ADNc de la protéine SR-p70a à partir de cellules HT-29 (Adénocarcinome de colon humain).

1) *Culture des cellules HT-29*

Les cellules sont cultivées en milieu McCoy 5 (GIBCO 28600-023) additionné de 10 % de sérum foetal de veau (GIBCO 10081-23) et 50 mg/l de gentamycine jusqu'à semi-confluence.

2) *Préparation de l'ADN complémentaire*

L'ARN messager est préparé comme décrit dans l'EXEMPLE I.2. L'ADNc est préparé de manière similaire à celle décrite dans l'EXEMPLE I.3 avec 5 µg d'ARN messager total en utilisant une amorce poly(T)₁₂. La réaction n'est pas interrompue avec de l'EDTA.

3) Amplification spécifique de l'ADNc humain par la technique dite de PCR

La polymérisation est réalisée avec 4 µl d'ADNc dans 50 µl final avec le tampon de composition suivante : Tris HCl 10 mM pH 8,3, MgCl₂ 2,5 mM, KCl 50 mM en présence de 10 % DMSO, dNTP 0,5 mM, 4 µg/ml de chacune des deux amores nucléiques et de 2,5 unités de Taq ADN polymérase (Boehringer). Les couples d'amores ont été choisis sur la séquence nucléique du clone SR-p70 de COS-3, notamment en amont de l'ATG d'initiation et en aval du TGA d'arrêt de traduction et sont de compositions suivantes :

10

amorce sens : ACT GGT ACC GCG AGC TGC CCT CGG AG
site de restriction Kpn I

15

amorce antisens : GAC TCT AGA GGT TCT GCA GGT GAC TCA G.
site de restriction Xba I

La réaction est réalisée durant 30 cycles 94°C/1 minute, 54-60°C/1 minute 30 secondes et 72°C/1 minute 30 secondes, suivi d'un dernier cycle de 72°C/6 minutes.

20

4) Obtention de la séquence de l'ADNc humain

Dans un premier temps, le produit de PCR est éliminé des oligonucléotides sur une colonne de sephacyr S400 puis déssalé par chromatographie d'exclusion sur une colonne de polyacrylamide P10 (Biorad référence 1504144). Les réactions de séquençage sont réalisées à l'aide du kit Applied Biosystems (référence 401628) avec des oligonucléotides spécifiques de l'ADNc. La séquence obtenue est très similaire à celle du SR-p70a de singe et la protéine déduite contient 636 acides aminés (Figure 6).

30

De manière similaire, d'autres séquences issues de lignées ou de tissus humains ont été obtenues pour la partie codante du SR-p70 humain, notamment à partir du poumon ou du pancréas. Les protéines déduites de ces séquences sont identiques à celles obtenues pour la lignée HT-29.

35

5) Clonage de l'ADNc humain dans le plasmide pCDNA3 (Invitrogen V 790-20)

Le produit PCR obtenu en 3) ainsi que le plasmide sont digérés par les deux enzymes de restriction Kpn I et Xba I puis purifiés après migration sur un gel d'agarose 1 % à l'aide du kit Geneclean (Bio 101 référencé 3105). Après ligation avec 100 ng d'insert

et 10 ng de vecteur et transformation (technique décrite dans l'EXAMPLE I.3.g et i, les clones recombinants sont vérifiés par séquençage à l'aide du kit Applied Biosystems cité ci-dessus.

5

EXAMPLE III

Clonage de l'ADNc du SR-p70 de souris à partir de cellules AtT-20 (tumeur hypophysaire)

10

1) Culture cellulaire de la lignée AtT-20

Les cellules sont cultivées dans du milieu Ham F10 (GIBCO 31550-023) additionné de 15 % de serum de cheval (GIBCO 26050-047) et de 2,5 % de serum foetal de veau (GIBCO 10081-073) et de 50 mg/l de gentamycine jusqu'à semi-confluence.

15

2) Préparation de la banque d'ADN complémentaire

La banque est réalisée comme décrit dans l'EXAMPLE I, 2) et 3) à partir des cellules cultivées ci-dessus.

20

3) Criblage de la banque

a) Préparation des membranes

Les clones de la banque sont étalés sur du milieu LB gélosé (boîtes de petri diamètre 150) revêtu de membranes Biodyne A (PALL référence BNNG 132). Après une nuit à 37°C, les clones sont transférés par contact sur de nouvelles membranes. Ces dernières sont traitées en les déposant sur du papier Whatman 3 mm imbibé des solutions suivantes : NaOH 0.5 N, NaCl 1.5 M pendant 5 minutes puis Tris HCl 0.5 M pH 8 , NaCl 1.5 M pendant 5 minutes. Après un traitement à la protéinase K dans le tampon suivant : Tris HCl 10 mM pH 8 , EDTA 10 mM, NaCl 50 mM, SDS 0.1 %, protéinase K 100 µg/ml pendant une heure à température ambiante, les membranes sont lavées abondamment dans du 2 x SSC (sodium citrate NaCl), séchées, puis incubées au four sous vide à 80°C pendant 20 minutes.

b) Préparation de la sonde

Sur la base de séquences des ADNc SR-p70 de singe et d'humain, une première séquence a été réalisée sur un fragment amplifié à partir d l'ARNm de la lignée AtT-20 comme décrit dans l'EXAMPLE II.3 et 4 avec les oligomères de compositions suivantes :

30

35

amorce sens : GCC ATG CCT GTC TAC AAG

amorce antisens : ACC AGC TGG TTG ACG GAG.

Sur la base de cette séquence, une sonde oligomérique spécifique de souris a été choisie et présente la composition suivante :

5 GAG CAT GTG ACC GAC ATT G.

100 ng de la sonde sont marqués en 3' avec 10 unités de Terminal Transférase (Pharmacia) et 100 µCi de dCTP $\alpha^{32}P$ 3000 Ci/mmol (Amersham référence PB 10205) dans 10 µl du tampon suivant : Tris HCl 30 mM pH 7.6 , acide cacodylique 140 mM, CoCl₂ 1 mM, DTT 0.1 mM pendant 15 minutes à 37°C. Les nucléotides radiomarqués non incorporés sont éliminés sur une colonne de polyacrylamide P10 (Biorad, référence 1504144). La sonde obtenue a une activité spécifique environ de 5.10⁸ dpm/µg.

c) Préhybridation et hybridation

15 Les membranes préparées en a) sont préhybridées 30 minutes à 42°C dans 6 x SSC, 5 x Denhart's, 0.1 % SDS puis hybridées quelques heures dans le même tampon additionné de la sonde préparée en b) à raison de 10⁶ dpm/ml.

d) Lavage et exposition des membranes

20 Les membranes sont lavées deux fois à température ambiante dans le tampon 2 x SSC/SDS 0.1 % puis une heure à 56°C en 6 x SSC/SDS 0.1 %. Les clones hybridés sont révélés avec des films KODAK XOMAT. Un clone positif contenant le SR-p70 de souris est sélectionné et dénommé ci-après SR-p70c.

4) Séquençage du SR-p70 de souris et analyse de la séquence

25 La séquence est obtenue à l'aide du kit Applied Biosystem (référence 401628). La séquence protéique déduite de l'ADNc SR-p70c de souris (Figure 7) présente une très forte homologie avec celles de l'humain et de singe excepté dans la partie N-terminale qui diverge fortement (voir Figure 9). A l'aide de la technique dite de PCR, de manière similaire à celle décrite dans l'EXEMPLE II.3 et 4, une seconde séquence 30 5' (issue de la même banque AtT-20) a été obtenue (Figure 8). La séquence protéique N-terminale déduite (séquence dénommée SR-p70a) est très similaire à celle déduite des ADNc SR-p70 humain et de singe (SR-p70a) (Figure 9). La lignée AtT-20 présente donc au moins deux transcripts SR-p70. Ces 2 derniers divergent dans la partie N-terminale par des épissages différents.

EXEMPLE IV

1) Production de protéine recombinante SR-p70 dans *E. coli*

a) Construction du plasmide d'expression

5 Elle consiste à mettre la partie -COOH terminale de la protéine SR-p70a de singe, depuis la valine en position 427 à l'histidine -COOH terminale en position 637, en fusion avec la glutathione S-transferase (GST) du vecteur plasmidique pGEX-4T-3 (Pharmacia référence 27-4583). Pour cela, l'insert correspondant de la SR-p70a (position 1434 à 2066) a été amplifié par PCR avec 10 ng de plasmide contenant 10 l'ADNc SR-p70a de singe. Les amores nucléiques sont de composition suivante :

amorce sens : TTT GGA TCC GTC AAC CAG CTG GTG GGC CAG

site de restriction BamHI

15 amorce antisens : AAA GTC GAC GTG GAT CTC GGC CTC C.
site Sal I

20 Le fragment obtenu ainsi que le vecteur sont digérés par les enzymes de restriction BamHI et Sal I et le clonage est réalisé comme décrit dans l'EXAMPLE II.5. Le clone sélectionné est appelé pG SR-p70.

b) Expression et purification de la protéine fusion GST-pSR-p70

Cette étape a été réalisée en utilisant le kit "bulk GST purification module" (Pharmacia Référence 27-4570-01).

25 De manière schématique, le clone recombinant a été mis en culture à 37°C dans un litre de milieu 2x YTA + ampicilline 100 µg/ml. A DO 0,8, l'expression est induite avec 0,5 mM d'IPTG pendant 2 heures à 37°C. Après centrifugation, le culot cellulaire est repris dans du PBS froid puis soniqué par ultrason. Après adjonction de 1 % triton X-100, on incube 30 minutes sous agitation à température ambiante. Après centrifugation à 12 000 g, 10 minutes à 4°C, on récupère le surnageant. La purification est ensuite réalisée sur une colonne de chromatographie d'affinité glutathion sepharose 4B. La fixation et le lavage sont réalisées en tampon PBS et l'élution est réalisée par compétition avec du glutathion réduit. La concentration finale est amenée à 300 µg/ml de protéine fusion.

2) *Production de protéine SR-p70a dans les cellules COS-3.*

Les cellules COS-3 sont transfectées avec de l'ADN plasmidique pSE1 dans lequel a été cloné l'ADNc de SR-p70a de singe (EXEMPLE I.1) ou avec de l'ADN plasmidique du vecteur pSE1 en tant que témoin par la technique du DEAE Dextran : les cellules COS-3 sont ensemençées à 5×10^5 cellules par boîte de 6 cm en milieu de culture contenant 5 % de sérum de bovin foetal (EXEMPLE I.1). Après culture, les cellules sont rincées avec du PBS. On ajoute 1 ml du mélange suivant : milieu contenant 6,5 µg d'ADN, 250 µg/ml de DEAE Dextran et 100 µM de chloroquine. Les cellules sont incubées à 37°C en 5 % CO₂ durant 4 à 5 heures. Le milieu est aspiré, on ajoute 2 ml de PBS additionné de 10 % DMSO et les cellules sont incubées pendant une minute en remuant légèrement les boîtes. Le milieu est à nouveau aspiré et les cellules sont rincées deux fois avec du PBS. Les cellules sont alors incubées à 37°C avec du milieu contenant 2 % de sérum de bovin foetal pendant la durée de l'expression qui est généralement de 3 jours.

La protéine SR-p70a est alors analysée comme décrit dans l'EXEMPLE VI par immunoempreinte.

EXEMPLE V

20

Préparation d'anticorps spécifiques

150 µg de protéines de l'échantillon préparé selon l'EXEMPLE IV ont été utilisés pour immuniser un lapin (mâle de 1,5 à 2 kg environ, New-Zealand). Les immunisations ont été effectuées tous les 15 jours selon le protocole décrit par Vaitukaitis, Methods in Enzymology, 1981, 73, 46. Pour la première injection, un volume de solution antigénique est émulsifié par un volume d'adjuvant complet de Freund (Sigma référence 4258). Cinq rappels ont été administrés en adjuvant incomplet de Freund (Sigma référence 5506).

30

EXEMPLE VI

Détection de la protéine SR-p70 "Western immunoblotting" (immunoempreinte)

1) Matériels utilisés pour l'immunoempreinte

a) Lignées cellulaires utilisées pour l'immunoempreinte.

35 Les lignées cellulaires suivantes ont été cultivées, comme décrit dans le catalogue «catalogue of cell lines and hybridomas, 7th edition, 1992» de l'ATCC (American Typ

Culture Collection) : COS-3, CV-1 (lignée de cellules de rein de singe), HT-29, U-373MG (glioblastome humain), MCF7 (adenocarcinome mammaire humain), SKNAS (neuroblastome humain cultivé dans les mêmes conditions que COS-3), SK-N-MC (neuroblastome humain), IMR-32 (neuroblastome humain), CHP212 (neuroblastome humain cultivé dans les mêmes conditions que CV-1), Saos-2 (ostéosarcome), SK-OV-3 (adénocarcinome d'ovaire) et SW 480 (adénocarcinome de colon humain).

5 b) Cellules COS-3 transfectées par l'ADNc SR-p70a.

Les cellules COS-3 ont été transfectées comme décrit dans l'EXEMPLE IV.2. En tant que témoin, les cellules ont été transfectées avec de l'ADN plasmidique pSE1 ne contenant pas l'ADNc recombinant SR-p70a.

10 2) *Préparation des échantillons protéiques à partir de culture cellulaire eucaryote ou de cellules transfectées.*

Après culture, les cellules sont lavées avec du PBS puis reprises dans un tampon RIPA (PBS avec 1 % NP40, 0,5 % sodium déoxycholate, 0,5 % SDS) complémenté avec 10 µg/ml RNAse A, 20µg/ml DNase 1, 2 µg/ml aprotinine, 0,5 µg/ml leupeptine, 0,7 µg/ml pepstatine et 170 µg/ml PMSF. Les cellules sont soniquées par ultrason à 4 °C et laissées 30 minutes à 4°C. Après microcentrifugation à 12 000 rpm, on récupère le surnageant. La concentration de protéine est mesurée par la méthode de Bradford.

15 20 3) "Western blotting"

25 5 ou 50 µg de protéines (50 µg pour les lignées cellulaires et 5 µg pour des cellules transfectées) sont mis dans 0,2 volume du tampon d'électrophorèse 6X suivant : Tris HCl 0,35 mM pH 6,8 10,3 % SDS, 38 % glycérol, 0,6 mM DTT, 0,012 % bleu de bromophénol. Les échantillons sont déposés et mis à migrer dans un gel SDS-PAGE 10 % (30/0,8 Bis) puis électrotransférés sur une membrane de nitrocellulose.

30 35 4) *Révélation par l'anticorps*

La membrane est incubée 30 minutes dans le tampon de blocage TBST (Tris HCl 10 mM pH 8, NaCl 150 mM, 0,2 % Tween 20) additionné de 5 % de lait (GIBCO- SKIM MILK) à température ambiante. La membrane est successivement mise en présence de l'anticorps anti-SR-p70 (α SR-p70) dans le même tampon 16 heures à 4°C, lavée 3 fois pendant 10 minutes avec du TBST, puis incubée une heure à 37°C avec un second anticorps anti-immunoglobuline de lapin couplé avec de la peroxydase (SIGMA A055). Après trois lavages de 15 minutes, la révélation est effectuée en utilisant le kit ECL (Amersham RPN2106) par chimioluminescence.

Parallèlement, les mêmes échantillons ont été révélés par un anticorps anti-p53 (α p53) (sigma BP5312) suivi d'un second anticorps anti-immunoglobine de souris.

5) *Figures et résultats.*

5 Figure 10 : Immunoempreinte de la protéine SR-p70

Figure 10a : Détection de la protéine recombinante SR-p70

- colonnes 1 et 3 : COS-3 transfectée par le vecteur pSE1.
- colonnes 2 et 4 : COS-3 transfectée par le plasmide pSE1 contenant l'ADNc du

10 SR-p70a.

- colonnes 1 et 2 : Révélation par l'anticorps anti-SR-p70 (α SR-p70).
- colonnes 3 et 4 : Révélation par l'anticorps anti-P53 (α p53).

Figure 10b : Détection de la protéine endogène SR-p70

15 - colonnes 1 : COS-3 ; 2 : CV-1 ; 3 : HT-29 ; 4 : U-373 MG ; 5 : MCF7 ; 6 : SK-NAS ; 7 : SK-N-MC ; 8 : IMR-32 ; 9 : CHP212 ; 10 : Saos-2 ; 11 : SK-OV-3 et 12 : SW480.

A : Révélation par l'anticorps α SR-p70.

B : Révélation par l'anticorps α p53.

20 L'anticorps α SR-p70 reconnaît spécifiquement les protéines recombinantes (Figure 10a) et endogènes (Figure 10b) et ne croise pas avec la p53. L'analyse de lignées cellulaires humaines ou de singe montre que la protéine SR-p70 comme la p53 est généralement faiblement détectable. Par contre, lorsqu'une accumulation de p53 existe, la SR-p70 devient elle aussi plus facilement détectable (Figure 10b). Une étude par RT-PCR de la distribution des transcrits SR-p70 montre que le gène est exprimé dans tous les types cellulaires testés.

EXAMPLE VII

30 Clonage du gène du SR-p70 et localisation chromosomique.

1) *Clonage du gène SR-p70*

La banque utilisée est une banque de cosmides, préparée avec de l'ADN génomique humain purifié de placenta, et commercialisée par Stratagène (référence 95 1202).

35 Le criblage du gène est réalisé comme décrit dans l'exemple III.3 avec un fragment d'ADN SR-p70 marqué au 32 P avec le kit BRL "Random Primers DNA Labelling

Systems" (référence 18187-013). Les tampons d'hybridation et de lavage sont additionnés de 50 % formamide. Le dernier lavage est réalisé en 0,1 x SSC/SDS 0,1 % à 60°C. De manière similaire, le gène SR-p70 a été isolé à partir d'une banque préparée avec de l'ADN génomique de la souris black C57.

5

Une analyse et un séquençage partiel des clones mettent en évidence la présence de 14 exons avec une structure proche de celle du gène p53, notamment dans la partie centrale où la taille et le positionnement des exons sont très conservés (Figure 12). Cette structure a été définie partiellement chez la souris et chez l'homme.

10

A titre d'exemple, les séquences génomiques humaines du 3' de l'intron 1, de l'exon 2, de l'intron 2, et du 5' de l'exon 3 sont présentées dans la figure 13.

2) Localisation chromosomique du gène SR-p70 chez l'homme

15

Elle a été réalisée avec de l'ADN du gène SR-70 humain en utilisant la technique décrite par R. Slim et al., Hum. Genet., 1991, 88, 21-26. Cinquante mitoses ont été analysées dont plus de 80% avaient des doubles spots localisés en 1p36 sur les deux chromosomes et plus particulièrement en 1p36.2 -1p36.3 (Figure 11). L'identification du chromosome 1 et son orientation sont basées sur l'hétérochromatine de la constrictio secondaire. Les images ont été faites sur un microscope Zeiss Axiophot, saisies par une caméra CCD refroidie LHESA et traitées par Optilab.

20

EXAMPLE VIII

25

A) Mise en évidence d'un ARNm codant pour une protéine SR-p70 humaine déduite présentant à la fois une extrémité N-terminale plus courte et une divergence.

1) Cultures des cellules IMR-32 (neuroblastome humain)

30

Les cellules ont été cultivées comme décrit dans le catalogue "catalogue of cell lines and hybridomas, 7th edition, 1992" de l'ATCC (American Type Culture Collection).

2) Préparation de l'ADNc

35

L'ARN est préparé comme décrit dans l'exemple I.2.a. L'ADNc est préparé de manière similaire à celle décrite dans l'exemple I.3 avec 5 µg d'ARN total dans un volume final de 20 µl en utilisant une amorce poly (T)₁₂ et avec des nucléotides froids. La réaction n'est pas interrompue avec de l'EDTA.

3) *Amplification spécifique de l'ADNc SR-P70 par la technique dite de PCR*

La polymérisation est réalisée avec 2 µl d'ADNc dans 50 µl final avec le tampon de composition suivant : Tris HCl 50 mM pH 9,2, 16 mM (NH₄)₂ SO₄, 1,75 mM MgCl₂, en présence de DMSO 10%, de NTP 0,4 mM, de 100 ng de chacune des deux amorces nucléiques et de 3,5 unités du mélange des Taq et PWO polymérasées (Boehringer Mannheim, réf. 1681 842).

Le couple d'amorce est de composition suivante :

10 amorce sens : AGGCCGGCGTGGGAAG (position 16 à 32, Figure 6)
amorce antisens : CTTGGCGATCTGGCAGTAG (position 503 à 485, Figure 6).

La réaction est réalisée durant 30 cycles à 95°C/30 secondes, 58°C/1 minute et 68°C/2 minutes 30 secondes suivi d'un dernier cycle de 68°C/10 minutes.

15 Le produit PCR est soumis à une électrophorèse sur un gel d'agarose 1% (tampon TAE). Après coloration au bromure d'éthidium, deux bandes majeures sont révélées : une bande d'une taille d'environ 490 pb (taille attendue (voir Figure 6)) et une bande supplémentaire d'une taille d'environ 700 pb. Cette dernière est extraite du gel à l'aide du kit "geneclean" (Bio 101, réf 1001 400). Après un dessalage sur une colonne de polyacrylamide P10 (Biorad, réf 15011050), le fragment est soumis à une nouvelle amplification par PCR durant 10 cycles comme décrit ci-dessus.

20 25 4) *Détermination de la séquence du produit amplifié*
Dans un premier temps, le produit PCR est éliminé des oligonucléotides sur une colonne de sephacyrl S400 (Pharmacia 17-0609-01) puis dessalé sur une colonne de P10. La réaction de séquençage est réalisée à l'aide du kit Applied Biosystems (réf. 401 628) (373 DNA sequencer) avec l'amorce antisens.

30 La séquence obtenue est identique à la séquence de l'ADNc SR-p70 (exemple II.4) avec une insertion de 198 pb entre les positions 217 et 218 (Figure 14). La séquence protéique N-terminale déduite (séquence dénommée SR-p70d) est plus courte de 49 acides aminés avec une divergence des 13 premiers acides aminés (séquence ID N°13). Il y a donc co-existence d'au moins deux transcrits différents SR-p70 comme déjà décrit pour la lignée AtT-20 de souris.

B) Clonag du SR-p70 humain et mise en évidenc d'un ARNm codant pour une protéine SR-p70 humaine déduite présentant la même extrémité N-terminale que le SR-p70d et une divergence dans la partie C-terminale

1) Amplification spécifique de l'ADNc du SR-p70 par la technique dite de PCR

5

L'amplification a été réalisée comme décrit dans l'EXEMPLE VIII.A à partir d'ARN purifié des cellules IMR-32 avec le couple d'amorces de composition suivante:

amorce sens : GCG GCC ACG ACC GTG AC (position 160 à 176, séquence ID N° 11)

10

amorce antisens : GGC AGC TTG GGT CTC TGG (position 1993 à 1976, Figure 6).

15

Après élimination de l'excés d'amorces sur colonne S400 et dessalage sur colonne P10, 1µl de l'échantillon est de nouveau soumis à une PCR avec le couple d'amorces de composition suivante :

amorce sens : TAT CTC GAG CTG TAC GTC GGT GAC CCC

Xho I

(position 263 à 280, séquence ID

N° 11)

amorce antisens : ATA TCT AGA TCA GTG GAT CTC GGC CTC

Xba I

(position 1943 à 1926, Figure 6).

20

2) Clonage du produit amplifié dans le plasmide pCDNA3

25

Le produit PCR obtenu en 1) est déssalé sur colonne P10, digéré par les enzymes de restriction Xho I et Xba I, puis cloné dans le plasmide pCDNA3 comme décrit dans l'EXEMPLE II.5 . Deux clones recombinants sont séquencés à l'aide du kit Applied Biosystems avec les oligonucléotides spécifiques de l'ADNc du SR-p70.

30

La première séquence obtenue correspond à la séquence complète de l'ARNm codant pour le SR-p70 décrit dans l'EXEMPLE VIII.a . La protéine déduite comporte 587 acides aminés (séquence ID N° 13 et Figure 16).

35

La seconde séquence obtenue est identique à la séquence de l' ADNc de SR-p70d décrite ci-dessus mais avec deux délétions de 149 pb et de 94 pb entre les positions 1049 et 1050 d'une part et entre les positions 1188 et 1189 d'autre part (séquence ID N° 14 et Figure 15). La séquence protéique déduite de cette seconde séquence révèle une protéine ayant une partie N-terminale plus courte de 49 acides aminés avec une divergence dans les 13 premiers acides aminés ainsi qu'un divergence de

séquence protéique entre les acides aminés 350 et 397 (séquence ID N° 15 et Figure 16) (séquence dénommée SR-p70e). La protéine déduite comporte 506 acides aminés.

- 5 C) Mise en évidence d'un ARNm codant pour une protéine SR-p70 humaine déduite présentant une extrémité N-terminale plus courte

10 1) Culture des cellules SK-N-SH (neuroblastome humain)

15 Les cellules sont cultivées comme décrit dans le « catalogue of cell lines and hybridomas, 7th edition, 1992 » de l'ATCC (American Type Culture Collection).

20 2) Préparation de l'ADNc et amplification de l'ADNc du SR-p70 par la technique dite de PCR

15

Ces étapes sont réalisées comme décrit dans l'EXEMPLE VIII.A avec le couple d'amorces de composition suivante:

25 amorce sens : AGG GGA CGC AGC GAA ACC (position 128 à 145, Figure 17)

20 amorce anti sens : GGC AGC TTG GGT CTC TGG (position 1993 à 1976, Figure 6).

30 Le séquençage est réalisé avec le kit Applied Biosystem avec des amorces spécifiques de l'ADNc du SR-p70 et révèle deux ADNc :

- un premier ADNc correspondant à l'ARNm codant pour le SR-p70a
- un second ADNc présentant une délétion de 98 pb entre les positions 24 et 25 (séquence ID N° 16 et Figure 15).

35 Cette délétion comprend l'ATG d'initiation de traduction du SR-p70a. La protéine déduite (dénommée SR-p70f) de ce second ADNc présente un ATG initiateur de traduction en aval correspondant à un ATG interne du SR-p70a. La protéine déduite comporte donc 588 acides aminés (séquence ID N° 17 et Figure 16) et est tronquée des 48 acides aminés N-terminaux du SR-p70a.

40 D) Mise en évidence d'un ARNm codant pour le SR-p70b humain.

35

1) Culture des cellules K562

Les cellules sont cultivées comme décrit dans le "catalogue of cell lines and hybridomas, 7 th édition, 1992" de l'ATCC (American Type Culture Collection).

- 5 2) Préparation de l'ADNc, amplification de l'ADNc du SR-p70 par la technique dite de PCR et séquençage.

Ces étapes sont réalisées comme décrit dans l'EXEMPLE VIII.C.

Le séquençage révèle deux ADNc :

- 10 Un premier ADNc correspondant à l'ARNm codant pour le SR-p70a et un second ADNc présentant une déletion de 94 pb entre les positions 1516 et 1517 (séquence ID N° 18 et Figure 15). La protéine déduite (dénommée SR-p70b) comporte 199 acides aminés et présente une séquence C-terminale tronquée de 137 acides aminés par rapport au SR-p70a avec les 4 derniers acides aminés divergents (séquence ID N° 19 et Figure 21).
- 15

Cet ADNc est similaire à celui décrit dans l'EXEMPLE I relatif au SR-p70b de singe.

- 20 Les molécules décrites dans cet exemple (EXEMPLE VIII. A. B. C. et D) mettent en évidence des variants du SR-p70 consécutifs à des épissages différentiels de l'ARNm primaire transcrits par le gène SR-p70.

- 25 Le SR-p70a est codé par un ARNm composé de 14 exons (voir EXEMPLE VII). C'est la protéine de référence. Le SR-p70b est consécutif à une insertion entre les exons 3 et 4 et à l'absence des exons 11 et 13. Le SR-p70f est consécutif à l'absence de l'exon 2. Cet exemple décrit l'existence de variants du SR-p70 de manière non exhaustive, avec une probabilité forte d'existence d'autres variants. De même, l'existence de ces variants décrits dans cet exemple ainsi que le SR-p70a ne se limite pas aux lignées dans lesquelles ils ont été mis en évidence. En effet des études effectuées par RT-PCR ont montré que ces variants sont retrouvés dans les diverses lignées étudiées.

- 30 De plus, la méthionine d'initiation du SR-p70f correspond à une méthionine interne du SR-p70a, suggérant la possibilité d'initiation en aval sur l'ARNm codant pour le SR-p70a.

EXEMPLE IX

Obtention d'une séquence 5' de l'ARNm SR-P70a humaine.

5

1) Amplification de l'extrémité de 5' de l'ADNc SR-P70 par PCR

La culture cellulaire et les préparations d'ARN total et d'ADNc sont réalisées comme décrit dans l'EXAMPLE VIII.1 et 2. La matrice ARN est hydrolysée par incubation 5 minutes à 65°C après adjonction de 4 µl EDTA 500 mM et 4 µl NaOH 2N. L'échantillon est ensuite dessalé sur colonne P10. L'ADNc est allongé en 3' avec une "queue" de dG comme décrit dans l'EXAMPLE I.3.d, dans un volume final de 40 µl. Après adjonction de 4 µl EDTA 500 mM et 4 µl NaOH 2N, l'ADNc est incubé à 65°C pendant 3 minutes puis dessalé sur une colonne P10. L'amplification par PCR est réalisée comme décrit dans l'EXAMPLE VIII.3 avec 8 µl d'ADNc et durant 30 cycles avec le couple d'amorces de composition suivante :

amorce sens : C C C C C C C C C C C N (où N est égal à G,A ou T)
amorce antisens : CCATCAGCTCCAGGCTCTC (position 1167 à 1149, Figure 6).

Après élimination de l'excès d'amorces sur colonne S400 et dessalage sur colonne P10, 1 µl de l'échantillon est de nouveau soumis à un PCR avec le couple de composition suivante :

amorce sens : C C C C C C C C C C C N
amorce antisens : CCAGGACAGGCGCAGATG (position 928 à 911, Figure 6).

L'échantillon, de nouveau passé sur une colonne S400 et une colonne P10, est soumis à une troisième amplification durant 20 cycles avec le couple suivant :

amorce sens : C C C C C C C C C C C N
amorce antisens : CTTGGCGATCTGGCAGTAG (position 503 à 485, Figure 6).

30

2) Détermination de la séquence 5' ADNc SR-P70

La séquence est réalisée comme décrit dans l'EXAMPLE VIII.4). Cette séquence fait apparaître un 5' non codant d'au moins 237 bases en amont de l'ATG d'initiation du SR-p70a (Figure 17). Par comparaison de cette séquence (obtenue à partir de la lignée IMR-32) avec celle obtenue à partir de la lignée HT-29 notamment (Figure 6), deux différences ponctuelles (Figure 17 : voir caractères gras) sont mises en évidence (G → A et C → T) respectivement positionné à -20 et -30 de l'ATG

d'initiation du SR-p70a (Figures 6 et 17). Cette variabilité est située au niveau de l'exon 2 (Figure 13). Il n'est pas exclu que cette variabilité se retrouve également à l'intérieur d'une phase codante consécutivement à un épissage alternatif comme décrit dans les EXEMPLES III chez la souris et VIII chez l'homme ou bien consécutivement à une initiation de la traduction sur un CTG (comme cela a été démontré pour le FGFb (Proc. Natl. Acad. Sci USA, 1989, 86, 1836 - 1840).
5 De même, il n'est pas exclu que cette variabilité ait une implication sur la traduction du SR-p70 ou sur l'épissage de l'ARN primaire.
En tout état de cause, cette variabilité, probablement d'origine allélique, peut servir de 10 marqueur soit au niveau génomique (voir EXEMPLE XI), soit au niveau de l'ARNm (voir EXEMPLE X).

EXEMPLE X

1) *Analyse par PCR, de l'expression transcriptionnelle du SR-P70a dans les 15 échantillons cellulaires (RT - PCR)*

Les cultures cellulaires (SK-N-AS, SK-N-MC, HT-29, U-373MG, SW480, IMR-32, CHP212) sont réalisées comme décrit dans l'exemple VI.1.a (références au catalogue "catalogue of cell lines and hybridomas, 7th edition" 1992 de l'ATCC).

20 La préparation de l'ADNc et l'amplification par PCR sont réalisées comme décrit dans l'exemple VIII.2 et 3. Le couple d'amorce utilisé est de composition suivante :

amorce sens : AGGGGACGCAGCGAAACC (position 128 à 145, Figure 17)

amorce antisens : GGCAGCTTGGGTCTCTGG (position 1993 à 1976, Figure 6).

25 Les échantillons sont analysés par électrophorèse sur un gel d'agarose 1% et révélation au Bromure d'éthidium (Figure 18).

La taille de la bande obtenue dans les échantillons correspond à la taille attendue (environ 2 kb, Figures 6 et 17). L'intensité des bandes obtenues est reproductible. 30 Une réamplification de 1 µl de l'échantillon dans les mêmes conditions durant 20 cycles fait apparaître une bande dans chacun des échantillons.

2) *Détermination de la séquence des produits amplifiés*

Après passage des échantillons sur colonnes S400 et P10, le séquençage est réalisé 35 sur le séquenceur 373 de Applied Biosystems avec le kit de référence 401 628. Les amorces utilisées sont entre autres les suivantes :

		position	figure
5	AGGGGACGCAGCGAAACC	128 à 145	22
	CTTGGCGATCTGGCAGTAG	503 à 485	6
	GATGAGGTGGCTGGCTGGA	677 à 659	6
	CCATCAGCTCCAGGCTCTC	1167 à 1149	6
	TGGTCAGGTTCTGCAGGTG	1605 à 1587	6
	GGCAGCTGGGTCTCTGG	1993 à 1976	6
10	Aucune différence protéique du SR-p70a n'a été décelée. Cependant, les séquences obtenues font apparaître une double variabilité aux positions -20 et -30 en amont de l'ATG d'initiation du SR-p70a (Figures 6 et 17). Cette variabilité, probablement d'origine allélique, permet de définir deux classes de transcrits : une première classe présentant un G à la position -30 et un C à la position -20 (classe G-30/C-20) et une seconde classe présentant une différence aux deux positions : un A en -30 et un T en -20 (classe A-30/T-20).		
	Première classe : SK-N-AS, SK-N-MC, HT-29, U-373MG, SW480.		
	Deuxième classe : IMR-32, CHP212.		

EXEMPLE XI

20

Méthode d'analyse pour la détermination de la répartition allélique du gène SR-p70 dans un échantillonnage de 10 personnes.

25 Cette répartition allélique est basée sur la variabilité allélique mise en évidence dans les exemples IX et X :

- Allèle G-30/C-20 présentant respectivement un G et un C aux positions -30 et -20 en amont de l'ATG d'initiation du SR-p70a.
 - Allèle A-30/T-20 présentant respectivement un A et un T aux mêmes positions.
- Cette variabilité peut être mise en évidence par l'utilisation d'enzymes de restriction différenciant les deux allèles (Figure 13). A titre d'exemple :
- Enzyme Bpl I présentant un site de coupure uniquement sur l'allèle G-30/C-20 dans la zone d'intérêt (ce site englobe les deux positions variables).
 - Enzyme StyI présentant un site de coupure uniquement sur l'allèle A-30/T-20 dans la zone d'intérêt.

35

1) *Amplification génomique de l'exon 2 par PCR*

La réaction de polymérisation est réalisée, avec 500 ng d'ADN génomique purifié, dans 50 µl final avec les conditions décrites dans l'exemple VIII.3.

Le couple d'amorces est de composition suivante :

5

amorce sens : CACCTACTCCAGGGATGC (position 1 à 18, Figure 13)
amorce antisens : AGGAAAATAGAAGCGTCAGTC (position 833 à 813, Figure 13).

10 La réaction est réalisée durant 30 cycles comme décrit dans l'EXEMPLE VIII.3.
Après élimination de l'excès d'amorce sur une colonne S400 et dessalage sur une colonne de P10, 1 µl de l'échantillon est amplifié de nouveau durant 25 cycles dans les mêmes conditions avec le couple d'amorces suivant :

15 amorce sens : CAGGCCCACTTGCGCTGCC (position 25 à 32, Figure 13)
amorce antisens : CTGTCCCCAAGCTGATGAG (position 506 à 488, Figure 13).

Les produits amplifiés sont soumis à une électrophorèse sur un gel d'agarose 1% (Figure 19-A).

20 2) *Digestion par l'enzyme de restriction StyI*

Les échantillons sont au préalable dessalés sur une colonne P10 puis digérés par l'enzyme de restriction StyI (BRL 15442-015) dans le tampon de composition suivant : Tris HCl pH 8, 50 mM, NaCl 100 mM, MgCl₂ 10 mM, à 37°C pendant 30 mn. Les produits de digestion sont analysés par électrophorèse sur un gel d'agarose 1% (tampon TAE). La révélation est réalisée par coloration au bromure d'ethidium (Figure 19-B).

30 Une bande de 482 paires de bases caractérise l'allèle G-30/C-20 (Figures 13 et 19). La présence d'une bande de 378 paires de bases et d'une bande de 106 paires de bases caractérisent l'allèle A-30/T-20 (allèle présentant un site de coupure StyI). Sur l'échantillonnage de 10 personnes, 2 personnes présentent les allèles G-30/C-20 et A-30/T-20, les 8 autres personnes étant homozygotes avec l'allèle G-30/C-20. L'étude d'un nouvel échantillonnage de 9 personnes a mis en évidence 3 personnes hétérozygotes présentant les allèles G-30/C-20 et A-30/T-20, les 6 autres personnes étant homozygotes pour l'allèle G-30/C-20.

35

EXEMPLE XII

5 Test de réversion de transformation de la lignée SK-N-AS par transfection avec l'ADNc SR-p70.

Le vecteur d'expression utilisé est décrit dans l'EXAMPLE II.5 et schématisé dans la figure 15. La méthode utilisée est celle dite du phosphate de calcium décrite par Graham et al. (Virology 1973, 54, 2, 536-539). La lignée est ensemencée à raison de 10 5.10^5 cellules par boîte de 6 cm de diamètre dans 5 ml du milieu décrit dans l'exemple I.1.. Les cellules sont mises en culture à 37°C et à 5% de CO₂ pendant une nuit. Le milieu de transfection est préparé de la manière suivante : le mélange suivant est réalisé en ajoutant dans l'ordre 1 ml de tampon HEBS (NaCl 8 mg/ml, KCl 370 µg/ml, Na₂HPO₄-2H₂O 125 µg/ml, Dextrose 1 mg/ml, Hepes pH 7,05 5 mg/ml), 10 µg 15 du plasmide à transfacter et 50 µl CaCl₂ 2,5 M ajouté goutte à goutte. Le milieu de transfection est laissé 30 mn à température ambiante puis ajouté goutte à goutte sur le milieu contenu dans la boîte de culture. Les cellules sont incubées 5 à 6 heures à 37° C/5% CO₂. Après aspiration du milieu, 5 ml de milieu frais contenant 2% de sérum du bovin foetal sont ajoutés. Après 48 heures à 37° C/5% CO₂, les cellules 20 sont rincées avec du PBS, décollées par trypsinisation, diluées dans 10 ml de milieu de culture (5% sérum de bovin foetal) et étalées dans une boîte de 10 cm de diamètre (la dilution peut être ajustée en fonction de l'efficacité de transfection). Après une nouvelle incubation durant 10 heures (le temps que les cellules adhèrent), les cellules 25 sont passées en sélection par adjonction de G418 à la concentration finale de 600 µg/ml équivalent génétidine durant 15 à 21 jours (le milieu est changé tous les jours). Les clones obtenus sont alors rincés au PBS, fixés à l'éthanol 70%, séchés, colorés avec 1 % de cristal violet, puis comptabilisés.

Quatre transfections plasmidiques ont été réalisées en double :

- 30 - plasmide pCDNA3 sans insert
- plasmide pCDNA3/SR-p70 contenant l'ADNc - SR-P70a humaine
- plasmide pCDNA3/SR-p70 Mut contenant l'ADNc - SR-p70a présentant une mutation à la position 293 AA (R → H) qui est analogue à la mutation 273 (R → H) dans le domaine de fixation à l'ADN de la p53
- témoin sans plasmide.

35

Le résultat est exprimé en nombre de clones par boîte.

	Expérience 1	Expérience 2	Moyenne
5	pCDNA3	172	353
	pCDNA3 / SR-p70	13	8
	pCDNA3 / SR-p70 mut	92	87
	Absence de plasmide	1	3
			2

Le nombre de clones obtenu par transfection avec le plasmide pCDNA3/SR-p70 est de 25 fois inférieur au nombre de clones obtenu avec le témoin pCDNA3 et de 9 fois inférieur au nombre de clones obtenu avec le pcDNA3/SR-p70 Mut, indiquant une mortalité ou un arrêt de division cellulaire des cellules transfectées par l'ADNc SR-p70. Ce résultat n'est pas la conséquence d'une toxicité au vue des clones obtenus avec l'ADNc SR-p70 muté mais probablement d'une apoptose comme cela a été démontré pour la protéine p53 (Koshland et al., Sciences, 1993, 262, 1953-1981).

EXEMPLE XIII

Rôle biologique de la protéine SR-p70.

L'homologie de structure entre le domaine de fixation à l'ADN de la p53 et la région centrale de la protéine SR-p70 permet d'inférer que la SR-p70 est un facteur de transcription (cf. Figures 1 et 2). En effet, la p53 (393 acides aminés) est constituée de plusieurs domaines fonctionnels. La région N-terminale (1-91 acides aminés) est impliquée dans l'activation de la transcription, et contient des sites d'interaction à différentes protéines cellulaires et virales. La partie centrale (acides aminés 92 à 292) permet la fixation aux séquences d'ADN spécifiques situés dans les régions promotrices de certains gènes (la majorité des mutations ponctuelles inactivant la p53 sont localisées dans cette région), elle présente également de nombreux sites d'interaction avec des protéines virales qui inhibent son activité. Enfin, les 100 derniers acides aminés de la p53 sont responsables de son oligomérisation ainsi que de la régulation de celle-ci (Hainaut P., Current Opinion in Oncology, 1995, 7, 76-82 ; Prokocimer M., Blood, 1994, 84 n°8, 2391-2411).

L'homologie de séquence entre p53 et SR-p70 est significative notamment en ce qui concerne les acides aminés impliqués directement dans l'interaction à l'ADN suggérant que la SR-p70 se fixe aux sites p53 sur l'ADN. Ces acides aminés

correspondent très exactement à ce qu'on appelle les "hot spot", acides aminés fréquemment mutés dans les tumeurs humaines (SWISS PROT : SW : P53_human et Prokocimer M., Blood, 1994, 84 n°8, 2391-2411). De cette homologie, on peut déduire que la protéine SR-p70 exerce un contrôle sur l'activité des gènes régulés par la p53, soit indépendamment de celle-ci soit en formant des hétérooligomères avec cette dernière.

En conséquence, à l'instar de la p53, les produits du gène SR-p70 doivent être impliqués dans le contrôle et la régulation du cycle cellulaire provoquant des arrêts du cycle (momentanés ou définitifs), et la mise en oeuvre de programmes tels que : la réparation de l'ADN, la différenciation ou la mort cellulaire. L'existence d'activités "p53-like" avait été fortement présentée avec la mise en évidence chez les souris p53^{-/-}, d'activités de réparation de l'ADN et de mort cellulaire en réponse aux radiations ionisantes (Strasser et al., Cell, 1994, 79, 329-339). Les auteurs de la présente invention ont localisé le gène SR-p70 humain dans la région téloïérique du bras court du chromosome 1; précisément en 1p36.2-36.3, la plus petite région délétée (SRO) commune à une majorité de neuroblastomes et d'autres types de tumeurs (mélanomes et carcinomes) (White et al., PNAS, 1995, 92, 5520-5524). Cette région de perte d'hétérozygotie (LOH) délimite le locus d'un gène suppresseur de tumeur dont la perte d'activité serait la cause de la formation des tumeurs. Il est important de rappeler que cette région est également sujette à "l'empreinte maternelle" ; l'allèle maternel est préférentiellement perdu dans les neuroblastomes présentant la délétion 1p36 (sans amplification de N-Myc) (Caron et al., Hum. Mol. Gen., 1995, 4, 535-539). Le gène sauvage SR-p70 introduit et exprimé dans des cellules de neuroblastome permet la réversion de leur transformation. La perte de cette activité anti-oncogénique est donc associée au développement de la tumeur. La région 1p36 présente une homologie syngénique avec le segment distal du chromosome 4 de souris. Dans cette région a été localisé le gène *curly tail* (*ct*) (Beier et al., Mammalian Genome, 1995, 6, 269-272) impliqué dans les malformations congénitales du tube neural (MTN : *spina bifida*, anencéphalie...). La souris *ct* est le meilleur modèle animal d'étude de ces malformations. Il est admis que ces malformations résultent d'anomalies de la prolifération cellulaire. Compte tenu de la nature du gène SR-p70 et de sa localisation chromosomique, une des hypothèses est que le SR-p70 pourrait être l'homologue humain de *ct* et qu'à ce titre, la détection des mutations précoces et les anomalies chromosomiques concernant ce gène devraient permettre par exempl comme application, l'identification des personnes à risques (0,5-1 % des nouveaux-nés atteints par MTN), et la mise en oeuvre de traitements

préventifs (Neumann *et al.*, Nature Genetics, 1994, 6, 357-362 ; Di Vinci *et al.*, Int. J. Cancer, 1994, 59, 422-426 ; Moll *et al.*, PNAS, 1995, 92, 4407-4411 ; Chen *et al.*, Development, 1995, 121, 681-691).

5

EXAMPLE XIV

Etude allélique du gène SR-p70.

- 10 Les allèles GC et AT sont identifiés aisément par restriction Styl des produits PCR de l'exon 2 (voir exemple XI). On a donc pu déterminer ainsi, chez des individus hétérozygotes GC/AT et porteurs de tumeurs neuroblastomes, l'allèle SR-p70 perdu (GC ou AT) et cela malgré la présence de tissu contaminant sain.
D'une façon surprenante, lorsque la même analyse est réalisée sur l'ARN, un seul allèle est mis en évidence indépendamment de présence ou non d'une délétion et plus surprenant encore, malgré la présence de tissu sain. Cela suggère que l'empreinte (expression différentielles des deux allèles) existerait également dans le tissu contaminant.
- 15 Pour le vérifier, on a répété la même analyse sur de l'ARN provenant des cellules sanguines d'individus sains hétérozygotes GC/AT. Un seul des deux types de transcrit a été détecté également dans ces cellules. Ce résultat confirme l'observation faite sur les échantillons tumoraux quant à l'existence d'une empreinte génétique généralisée pour le gène SR-p70.
- 20 Les implications de cette découverte sont importantes puisqu'elle permet de postuler qu'une seule mutation sporadique inactivant l'allèle actif SR-p70 entraînera une perte d'activité et cela potentiellement dans tous les tissus.
- 25 L'absence de données précises sur la fonction biologique SR-p70 ne permet pas de mesurer les conséquences de cette perte d'activité SR-p70 pour la cellule. Néanmoins, sa forte homologie avec la protéine suppresseur de tumeur p53, ainsi que la démonstration que la SR-p70 est un facteur de transcription capable d'utiliser le promoteur P21^{WT}, suggère un rôle de cette protéine dans le contrôle du cycle cellulaire et dans la différenciation.
- 30 Knudson and Meadows 1980 (New Eng. J. Med. 302 :1254-56) considèrent les neuroblastomes IV-S comme une collection de cellules non malignes de la crête neurale portant une mutation qui interfère avec leur différenciation normal .
- 35

Il est concevable que la pert d'activité SR-p70 tout comme la perte du contrôle p53 sur le cycle cellulaire favorise l'apparition d'anomalies cellulaires telles que l'aneuploidie, l'amplification (décris dans le cas des neuroblastomes) et d'autres remaniements génétiques pouvant provoquer la transformation cellulaire (Livingstone et al. 1992, Cell 71 :923-25 ; Yin et al. 1992, Cell 72 :937-48 ; Cross et al. 1995, Science 267 :1353-56 ; Fukasawa et al. 1996, science 271 :1744-47). Les neuroblastomes pourraient donc provenir à leur origine d'une perte d'activité temporaire ou définitive de la SR-p70, favorisant ainsi l'apparition d'événements oncogéniques et donc la progression tumorale.

10 Dans le cas de la délétion constitutionnelle 1p36 décrite par Biegel et al. 1993 (Am. J. Hum. Genet. 52 :176-82), il y a bien apparition de neuroblastome IV-S, et le gène concerné est NBS-1 (SR-p70).

15 En conclusion, ce qui est décrit pour les neuroblastomes pourrait également s'appliquer à d'autres types de tumeurs notamment ceux associés à des remaniements de l'extrémité du bras court du chromosome 1 (report 2 international workshop on human chr 1 mapping 1995, Cytogenetics and Cell Genet. 72 :113-154). Sur un plan thérapeutique, l'implication de la SR-p70 dans l'apparition de tumeurs devrait conduire à éviter l'utilisation d'agents mutagènes en chimiothérapie, compte tenu des risques de transformation cellulaire par ces produits, et leur préférer des substances non mutagènes qui stimulent la différenciation.

20 25 D'autre part, la fréquence d'apparition des allèles GC et AT est la suivante : dans la population, Fréquence(AT)=0.15 et sur un échantillon de 25 patients (neuroblastomes), F(AT)=0.30. Ces statistiques indiquent que l'allèle AT pourrait être un facteur de prédisposition.

LISTE DE SEQUENCES

(1) INFORMATION GENERALE:

(i) DEPOSANT:

- (A) NOM: sanofi
- (B) RUE: 32-34 rue Marbeuf
- (C) VILLE: PARIS
- (E) PAYS: FRANCE
- (F) CODE POSTAL: 75008
- (G) TELEPHONE: 01 53 77 40 00
- (H) TELECOPIE: 01 53 77 41 33

(ii) TITRE DE L' INVENTION: SR-p70

(iii) NOMBRE DE SEQUENCES: 40

(iv) FORME LISIBLE PAR ORDINATEUR:

- (A) TYPE DE SUPPORT: Floppy disk
- (B) ORDINATEUR: IBM PC compatible
- (C) SYSTEME D' EXPLOITATION: PC-DOS/MS-DOS
- (D) LOGICIEL: PatentIn Release #1.0, Version #1.25 (OEB)

(2) INFORMATION POUR LA SEQ ID NO: 1:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 2874 paires de bases
- (B) TYPE: acide nucléique
- (C) NOMBRE DE BRINS: double
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN

(vi) ORIGINE:

- (A) ORGANISME: Cebus apella

(ix) CARACTERISTIQUE ADDITIONNELLE:

- (A) NOM/CLE: CDS
- (B) EMPLACEMENT: 156..2066

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 1:

TGCCTCCCCG CCCGCGCACC CGCCCCGAGG CCTGTGCTCC TGCGAAGGGG ACGCAGCGAA	60
GCCGGGGCCC GCGCCAGGCC GGCCGGGACG GACGCCGATG CCGGGAGCTG CGACGGCTGC	120
AGAGCGAGCT GCCCTCGGAG GCCGGTGTGA GGAAG ATG GCC CAG TCC ACC ACC Met Ala Gln Ser Thr Thr	173
1 5	
ACC TCC CCC GAT GGG GGC ACC ACG TTT GAG CAC CTC TGG AGC TCT CTG Thr Ser Pro Asp Gly Gly Thr Thr Phe Glu His Leu Trp Ser Ser Leu	221
10 15 20	
GAA CCA GAC AGC ACC TAC TTC GAC CTT CCC CAG TCA AGC CGG GGG AAT Glu Pro Asp Ser Thr Tyr Phe Asp Leu Pro Gln Ser Ser Arg Gly Asn	269
25 30 35	
AAT GAG GTG GTG GGT GGC ACG GAT TCC AGC ATG GAC GTC TTC CAC CTA Asn Glu Val Val Gly Gly Thr Asp Ser Ser Met Asp Val Phe His Leu	317
40 45 50	
GAG GGC ATG ACC ACA TCT GTC ATG GCC CAG TTC AAT TTG CTG AGC AGC Glu Gly Met Thr Thr Ser Val Met Ala Gln Phe Asn Leu Leu Ser Ser	365
55 60 65 70	
ACC ATG GAC CAG ATG AGC AGC CGC GCT GCC TCG GCC AGC CCG TAC ACC Thr Met Asp Gln Met Ser Ser Arg Ala Ala Ser Ala Ser Pro Tyr Thr	413
75 80 85	

CCG GAG CAC GCC GCC AGC GTG CCC ACC CAT TCA CCC TAC GCA CAG CCC Pro Glu His Ala Ala Ser Val Pro Thr His Ser Pro Tyr Ala Gln Pro 90 95 100	461
AGC TCC ACC TTC GAC ACC ATG TCG CCC GCG CCT GTC ATC CCC TCC AAC Ser Ser Thr Phe Asp Thr Met Ser Pro Ala Pro Val Ile Pro Ser Asn 105 110 115	509
ACC GAC TAT CCC GGA CCC CAC CAC TTC GAG GTC ACT TTC CAG CAG TCC Thr Asp Tyr Pro Gly Pro His His Phe Glu Val Thr Phe Gln Gln Ser 120 125 130	557
AGC ACG GCC AAG TCA GCC ACC TGG ACG TAC TCC CCA CTC TTG AAG AAA Ser Thr Ala Lys Ser Ala Thr Trp Thr Tyr Ser Pro Leu Leu Lys Lys 135 140 145 150	605
CTC TAC TGC CAG ATC GCC AAG ACA TGC CCC ATC CAG ATC AAG GTG TCC Leu Tyr Cys Gln Ile Ala Lys Thr Cys Pro Ile Gln Ile Lys Val Ser 155 160 165	653
GCC CCA CCG CCC CCG GGC ACC GCC ATC CGG GCC ATG CCT GTC TAC AAG Ala Pro Pro Pro Gly Thr Ala Ile Arg Ala Met Pro Val Tyr Lys 170 175 180	701
AAG GCG GAG CAC GTG ACC GAC ATC GTG AAG CGC TGC CCC AAC CAC GAG Lys Ala Glu His Val Thr Asp Ile Val Lys Arg Cys Pro Asn His Glu 185 190 195	749
CTC GGG AGG GAC TTC AAC GAA GGA CAG TCT GCC CCA GCC AGC CAC CTC Leu Gly Arg Asp Phe Asn Glu Gly Gln Ser Ala Pro Ala Ser His Leu 200 205 210	797
ATC CGT GTG GAA GGC AAT AAT CTC TCG CAG TAT GTG GAC GAC CCT GTC Ile Arg Val Glu Gly Asn Asn Leu Ser Gln Tyr Val Asp Asp Pro Val 215 220 225 230	845
ACC GGC AGG CAG AGC GTC GTG GTG CCC TAT GAG CCA CCA CAG GTG GGG Thr Gly Arg Gln Ser Val Val Val Pro Tyr Glu Pro Pro Gln Val Gly 235 240 245	893
ACA GAA TTC ACC ACC ATC CTG TAC AAC TTC ATG TGT AAC AGC AGC TGT Thr Glu Phe Thr Thr Ile Leu Tyr Asn Phe Met Cys Asn Ser Ser Cys 250 255 260	941
GTG GGG GGC ATG AAC CGA CGG CCC ATC CTC ATC ATC ACC CTG GAG Val Gly Gly Met Asn Arg Arg Pro Ile Leu Ile Ile Ile Thr Leu Glu 265 270 275	989
ACG CGG GAT GGG CAG GTG CTG GGC CGC CGG TCC TTC GAG GGC CGC ATC Thr Arg Asp Gly Gln Val Leu Gly Arg Arg Ser Phe Glu Gly Arg Ile 280 285 290	1037
TGC GCC TGT CCT GGC CGC GAC CGA AAA GCC GAT GAG GAC CAC TAC CGG Cys Ala Cys Pro Gly Arg Asp Arg Lys Ala Asp Glu Asp His Tyr Arg 295 300 305 310	1085
GAG CAG CAG GCC TTG AAT GAG AGC TCC GCC AAG AAC GGG GCT GCC AGC Glu Gln Gln Ala Leu Asn Glu Ser Ser Ala Lys Asn Gly Ala Ala Ser 315 320 325	1133
AAG CGC GCC TTC AAG CAG AGT CCC CCT GCC GTC CCC GCC CTG GGC CGG Lys Arg Ala Phe Lys Gln Ser Pro Pro Ala Val Pro Ala Leu Gly Pro 330 335 340	1181
GGT GTG AAG AAG CGG CGG CAC GGA GAC GAG GAC ACG TAC TAC CTG CAG Gly Val Lys Lys Arg Arg His Gly Asp Glu Asp Thr Tyr Tyr Leu Gln 345 350 355	1229
GTG CGA GGC CGC GAG AAC TTC GAG ATC CTG ATG AAG CTG AAG GAG AGC Val Arg Gly Arg Glu Asn Phe Glu Ile Leu Met Lys Leu Lys Glu Ser 360 365 370	1277

CTG GAG CTG ATG GAG TTG GTG CCG CAG CCG CTG GTA GAC TCC TAT CGG Leu Glu Leu Met Glu Leu Val Pro Gln Pro Leu Val Asp Ser Tyr Arg 375 380 385 390	1325
CAG CAG CAG CAG CTC CTA CAG AGG CCG AGT CAC CTA CAG CCC CCA TCC Gln Gln Gln Leu Leu Gln Arg Pro Ser His Leu Gln Pro Pro Ser 395 400 405	1373
TAC GGG CCG GTC CTC TCG CCC ATG AAC AAG GTG CAC GGG GGC GTG AAC Tyr Gly Pro Val Leu Ser Pro Met Asn Lys Val His Gly Gly Val Asn 410 415 420	1421
AAG CTG CCC TCC GTC AAC CAG CTG GTG GGC CAG CCT CCC CCG CAC AGC Lys Leu Pro Ser Val Asn Gln Leu Val Gly Gln Pro Pro Pro His Ser 425 430 435	1469
TCG GCA GCT ACA CCC AAC CTG GGA CCT GTG GGC TCT GGG ATG CTC AAC Ser Ala Ala Thr Pro Asn Leu Gly Pro Val Gly Ser Gly Met Leu Asn 440 445 450	1517
AAC CAC GGC CAC GCA GTG CCA GCC AAC AGC GAG ATG ACC AGC AGC CAC Asn His Gly His Ala Val Pro Ala Asn Ser Glu Met Thr Ser Ser His 455 460 465 470	1565
GGC ACC CAG TCC ATG GTC TCG GGG TCC CAC TGC ACT CCG CCA CCC CCC Gly Thr Gln Ser Met Val Ser Gly Ser His Cys Thr Pro Pro Pro Pro 475 480 485	1613
TAC CAC GCC GAC CCC AGC CTC GTC AGT TTT TTA ACA GGA TTG GGG TGT Tyr His Ala Asp Pro Ser Leu Val Ser Phe Leu Thr Gly Leu Gly Cys 490 495 500	1661
CCA AAC TGC ATC GAG TAT TTC ACG TCC CAG GGG TTA CAG AGC ATT TAC Pro Asn Cys Ile Glu Tyr Phe Thr Ser Gln Gly Leu Gln Ser Ile Tyr 505 510 515	1709
CAC CTG CAG AAC CTG ACC ATC GAG GAC CTG GGG GCC CTG AAG ATC CCC His Leu Gln Asn Leu Thr Ile Glu Asp Leu Gly Ala Leu Lys Ile Pro 520 525 530	1757
GAG CAG TAT CGC ATG ACC ATC TGG CGG GGC CTG CAG GAC CTG AAG CAG Glu Gln Tyr Arg Met Thr Ile Trp Arg Gly Leu Gln Asp Leu Lys Gln 535 540 545 550	1805
GGC CAC GAC TAC GGC GCC GCC GCG CAG CAG CTG CTC CGC TCC AGC AAC Gly His Asp Tyr Gly Ala Ala Ala Gln Gln Leu Leu Arg Ser Ser Asn 555 560 565	1853
GCG GCC GCC ATT TCC ATC GGC GGC TCC GGG GAG CTG CAG CGC CAG CGG Ala Ala Ala Ile Ser Ile Gly Ser Gly Glu Leu Gln Arg Gln Arg 570 575 580	1901
GTC ATG GAG GCC GTG CAC TTC CGC GTG CGC CAC ACC ATC ACC ATC CCC Val Met Glu Ala Val His Phe Arg Val Arg His Thr Ile Thr Ile Pro 585 590 595	1949
AAC CGC GGC CCC GGC GCC GGC CCC GAC GAG TGG GCG GAC TTC GGC Asn Arg Gly Gly Pro Gly Ala Gly Pro Asp Glu Trp Ala Asp Phe Gly 600 605 610	1997
TTC GAC CTG CCC GAC TGC AAG GCC CGC AAG CAG CCC ATC AAG GAG GAG Phe Asp Leu Pro Asp Cys Lys Ala Arg Lys Gln Pro Ile Lys Glu Glu 615 620 625 630	2045
TTC ACG GAG GCC GAG ATC CAC TGAGGGGCCG GGCCCCAGCCA GAGCCTGTGC Phe Thr Glu Ala Glu Ile His 635	2096
CACCGCCCAG AGACCCAGGC CGCCTCGCTC TCCTTCCTGT GTCCAAAAGT GCCTCCGGAG GCAGGGCCCTC CAGGCTGTGC CCAGGGAAAG GCAAGGTCCG GCCCCATGCCG CGGCACCTCA	2156 2216

CCGGCCCCAG GAGAGGCCA GCCACCAAAG CCGCCTGGG ACAGCCTGAG TCACCTGCAG	2276
AACCTTCTGG AGCTGCCCTA ATGCTGGGCT TGCGGGCAG GGGCCGGCCC ACTCTCAGCC	2336
CTGCCACTGC CGGGCGTGCT CCATGGCAGG CGTGGGTGGG GACCGCAGTG TCAGCTCCGA	2396
CCTCCAGGCC TCATCCTAGA GACTCTGTCA TCTGCCGATC AAGCAAGGTC CTTCCAGAGG	2456
AAAGAACCTT CTTCGCTGGT GGACTGCCAA AAAGTATTTC GCGACATCTT TTGGTCTGG	2516
AGAGTGGTGA GCAGCCAAGC GACTGTGTCT GAAACACCCT GCATTTTCAG GGAATGTCCC	2576
TAACGGGCTG GGGACTCTCT CTGCTGGACT TGGGAGTGGC CTTTGGCCCC AGCACACTGT	2636
ATTCTGCGGG ACCGCCTCCT TCCTGCCCT ACAACCACC AAAGTGTG TGAAATTGGA	2696
GAAAACGGG GAAGGCGCAA CCCCTCCAG GTGCGGGAAAG CATCTGGTAC CGCCTCGGCC	2756
AGTGCCTCTC AGCCTGGCCA CAGTCACCTC TCCTGGGAA ACCCTGGCA GAAAGGGACA	2816
GCCTGTCCCTT AGAGGACCGG AAATTGTCAA TATTTGATAA AATGATAACCC TTTCTAC	2874

(2) INFORMATION POUR LA SEQ ID NO: 2:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 637 acides aminés
- (B) TYPE: acide aminé
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: protéine

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 2:

Met Ala Gln Ser Thr Thr Ser Pro Asp Gly Gly Thr Thr Phe Glu			
1	5	10	15
His Leu Trp Ser Ser Leu Glu Pro Asp Ser Thr Tyr Phe Asp Leu Pro			
20	25	30	
Gln Ser Ser Arg Gly Asn Asn Glu Val Val Gly Gly Thr Asp Ser Ser			
35	40	45	
Met Asp Val Phe His Leu Glu Gly Met Thr Thr Ser Val Met Ala Gln			
50	55	60	
Phe Asn Leu Leu Ser Ser Thr Met Asp Gln Met Ser Ser Arg Ala Ala			
65	70	75	80
Ser Ala Ser Pro Tyr Thr Pro Glu His Ala Ala Ser Val Pro Thr His			
85	90	95	
Ser Pro Tyr Ala Gln Pro Ser Ser Thr Phe Asp Thr Met Ser Pro Ala			
100	105	110	
Pro Val Ile Pro Ser Asn Thr Asp Tyr Pro Gly Pro His His Phe Glu			
115	120	125	
Val Thr Phe Gln Gln Ser Ser Thr Ala Lys Ser Ala Thr Trp Thr Tyr			
130	135	140	
Ser Pro Leu Leu Lys Lys Leu Tyr Cys Gln Ile Ala Lys Thr Cys Pro			
145	150	155	160
Ile Gln Ile Lys Val Ser Ala Pro Pro Pro Gly Thr Ala Ile Arg			
165	170	175	
Ala Met Pro Val Tyr Lys Lys Ala Glu His Val Thr Asp Ile Val Lys			
180	185	190	
Arg Cys Pro Asn His Glu Leu Gly Arg Asp Phe Asn Glu Gly Gln Ser			

195	200	205
Ala Pro Ala Ser His Leu Ile Arg Val Glu Gly Asn Asn Leu Ser Gln		
210	215	220
Tyr Val Asp Asp Pro Val Thr Gly Arg Gln Ser Val Val Val Pro Tyr		
225	230	235
240		
Glu Pro Pro Gln Val Gly Thr Glu Phe Thr Thr Ile Leu Tyr Asn Phe		
245	250	255
Met Cys Asn Ser Ser Cys Val Gly Gly Met Asn Arg Arg Pro Ile Leu		
260	265	270
Ile Ile Ile Thr Leu Glu Thr Arg Asp Gly Gln Val Leu Gly Arg Arg		
275	280	285
Ser Phe Glu Gly Arg Ile Cys Ala Cys Pro Gly Arg Asp Arg Lys Ala		
290	295	300
Asp Glu Asp His Tyr Arg Glu Gln Gln Ala Leu Asn Glu Ser Ser Ala		
305	310	315
320		
Lys Asn Gly Ala Ala Ser Lys Arg Ala Phe Lys Gln Ser Pro Pro Ala		
325	330	335
Val Pro Ala Leu Gly Pro Gly Val Lys Lys Arg Arg His Gly Asp Glu		
340	345	350
Asp Thr Tyr Tyr Leu Gln Val Arg Gly Arg Glu Asn Phe Glu Ile Leu		
355	360	365
Met Lys Leu Lys Glu Ser Leu Glu Leu Met Glu Leu Val Pro Gln Pro		
370	375	380
Leu Val Asp Ser Tyr Arg Gln Gln Gln Leu Leu Gln Arg Pro Ser		
385	390	395
400		
His Leu Gln Pro Pro Ser Tyr Gly Pro Val Leu Ser Pro Met Asn Lys		
405	410	415
Val His Gly Gly Val Asn Lys Leu Pro Ser Val Asn Gln Leu Val Gly		
420	425	430
Gln Pro Pro Pro His Ser Ser Ala Ala Thr Pro Asn Leu Gly Pro Val		
435	440	445
Gly Ser Gly Met Leu Asn Asn His Gly His Ala Val Pro Ala Asn Ser		
450	455	460
Glu Met Thr Ser Ser His Gly Thr Gln Ser Met Val Ser Gly Ser His		
465	470	475
480		
Cys Thr Pro Pro Pro Tyr His Ala Asp Pro Ser Leu Val Ser Phe		
485	490	495
Leu Thr Gly Leu Gly Cys Pro Asn Cys Ile Glu Tyr Phe Thr Ser Gln		
500	505	510
Gly Leu Gln Ser Ile Tyr His Leu Gln Asn Leu Thr Ile Glu Asp Leu		
515	520	525
Gly Ala Leu Lys Ile Pro Glu Gln Tyr Arg Met Thr Ile Trp Arg Gly		
530	535	540
Leu Gln Asp Leu Lys Gln Gly His Asp Tyr Gly Ala Ala Ala Gln Gln		
545	550	555
560		
Leu Leu Arg Ser Ser Asn Ala Ala Ile Ser Ile Gly Gly Ser Gly		
565	570	575
Glu Leu Gln Arg Gln Arg Val Met Glu Ala Val His Phe Arg Val Arg		

580	585	590
His Thr Ile Thr Ile Pro Asn Arg Gly Gly Pro Gly Ala Gly Pro Asp		
595	600	605
Glu Trp Ala Asp Phe Gly Phe Asp Leu Pro Asp Cys Lys Ala Arg Lys		
610	615	620
Gln Pro Ile Lys Glu Glu Phe Thr Glu Ala Glu Ile His		
625	630	635

(2) INFORMATION POUR LA SEQ ID NO: 3:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 (A) LONGUEUR: 2034 paires de bases
 (B) TYPE: acide nucléique
 (C) NOMBRE DE BRINS: double
 (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc

(vi) ORIGINE:
 (A) ORGANISME: Cebus apella

(ix) CARACTERISTIQUE ADDITIONNELLE:
 (A) NOM/CLE: CDS
 (B) EMPLACEMENT: 156..1652

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 3:

TGCCTCCCCG CCCGCGCACC CGCCCCGAGG CCTGTGCTCC TGGAAAGGGG ACGCAGCGAA	60
GCCGGGGCCC GCGCCAGGCC GGCGGGACG GACGCCGATG CCCGGAGCTG CGACGGCTGC	120
AGAGCGAGCT GCCCTGGAG GCCGGTGTGA GGAAG ATG GCC CAG TCC ACC ACC Met Ala Gln Ser Thr Thr	173
1 5	
ACC TCC CCC GAT GGG GGC ACC ACG TTT GAG CAC CTC TGG AGC TCT CTG Thr Ser Pro Asp Gly Gly Thr Thr Phe Glu His Leu Trp Ser Ser Leu	221
10 15 20	
GAA CCA GAC ACC ACC TAC TTC GAC CTT CCC CAG TCA AGC CGG GGG AAT Glu Pro Asp Ser Thr Tyr Phe Asp Leu Pro Gln Ser Ser Arg Gly Asn	269
25 30 35	
AAT GAG GTG GTG GGT GGC ACG GAT TCC AGC ATG GAC GTC TTC CAC CTA Asn Glu Val Val Gly Gly Thr Asp Ser Ser Met Asp Val Phe His Leu	317
40 45 50	
GAG GGC ATG ACC ACA TCT GTC ATG GCC CAG TTC AAT TTG CTG AGC AGC Glu Gly Met Thr Thr Ser Val Met Ala Gln Phe Asn Leu Leu Ser Ser	365
55 60 65 70	
ACC ATG GAC CAG ATG AGC AGC CGC GCT GCC TCG GCC AGC CCG TAC ACC Thr Met Asp Gln Met Ser Ser Arg Ala Ala Ser Ala Ser Pro Tyr Thr	413
75 80 85	
CCG GAG CAC GCC GCC AGC GTG CCC ACC CAT TCA CCC TAC GCA CAG CCC Pro Glu His Ala Ala Ser Val Pro Thr His Ser Pro Tyr Ala Gln Pro	461
90 95 100	
AGC TCC ACC TTC GAC ACC ATG TCG CCC GCG CCT GTC ATC CCC TCC AAC Ser Ser Thr Phe Asp Thr Met Ser Pro Ala Pro Val Ile Pro Ser Asn	509
105 110 115	
ACC GAC TAT CCC GGA CCC CAC CAC TTC GAG GTC ACT TTC CAG CAG TCC Thr Asp Tyr Pro Gly Pro His His Phe Glu Val Thr Phe Gln Gln Ser	557
120 125 130	
AGC ACG GCC AAG TCA GCC ACC TGG ACG TAC TCC CCA CTC TTG AAG AAA	605

Ser Thr Ala Lys Ser Ala Thr Trp Thr Tyr Ser Pro Leu Leu Lys Lys 135 140 145 150	
CTC TAC TGC CAG ATC GCC AAG ACA TGC CCC ATC CAG ATC AAG GTG TCC Leu Tyr Cys Gln Ile Ala Lys Thr Cys Pro Ile Gln Ile Lys Val Ser 155 160 165	653
GCC CCA CCG CCC CCG GGC ACC GCC ATC CGG GCC ATG CCT GTC TAC AAG Ala Pro Pro Pro Gly Thr Ala Ile Arg Ala Met Pro Val Tyr Lys 170 175 180	701
AAG GCG GAG CAC GTG ACC GAC ATC GTG AAG CGC TGC CCC AAC CAC GAG Lys Ala Glu His Val Thr Asp Ile Val Lys Arg Cys Pro Asn His Glu 185 190 195	749
CTC GGG AGG GAC TTC AAC GAA GGA CAG TCT GCC CCA GCC AGC CAC CTC Leu Gly Arg Asp Phe Asn Glu Gly Gln Ser Ala Pro Ala Ser His Leu 200 205 210	797
ATC CGT GTG GAA GGC AAT AAT CTC TCG CAG TAT GTG GAC GAC CCT GTC Ile Arg Val Glu Gly Asn Asn Leu Ser Gln Tyr Val Asp Asp Pro Val 215 220 225 230	845
ACC GGC AGG CAG AGC GTC GTG GTG CCC TAT GAG CCA CCA CAG GTG GGG Thr Gly Arg Gln Ser Val Val Val Pro Tyr Glu Pro Pro Gln Val Gly 235 240 245	893
ACA GAA TTC ACC ACC ATC CTG TAC AAC TTC ATG TGT AAC AGC AGC TGT Thr Glu Phe Thr Thr Ile Leu Tyr Asn Phe Met Cys Asn Ser Ser Cys 250 255 260	941
GTG GGG GGC ATG AAC CGA CGG CCC ATC CTC ATC ATC ATC ACC CTG GAG Val Gly Met Asn Arg Arg Pro Ile Leu Ile Ile Ile Thr Leu Glu 265 270 275	989
ACG CGG GAT GGG CAG GTG CTG GGC CGC CGG TCC TTC GAG GGC CGC ATC Thr Arg Asp Gly Gln Val Leu Gly Arg Arg Ser Phe Glu Gly Arg Ile 280 285 290	1037
TGC GCC TGT CCT GGC CGC GAC CGA AAA GCC GAT GAG GAC CAC TAC CGG Cys Ala Cys Pro Gly Arg Asp Arg Lys Ala Asp Glu Asp His Tyr Arg 295 300 305 310	1085
GAG CAG CAG GCC TTG AAT GAG AGC TCC GCC AAG AAC GGG GCT GCC AGC Glu Gln Gln Ala Leu Asn Glu Ser Ser Ala Lys Asn Gly Ala Ala Ser 315 320 325	1133
AAG CGC GCC TTC AAG CAG AGT CCC CCT GCC GTC CCC GCC CTG GGC CCG Lys Arg Ala Phe Lys Gln Ser Pro Pro Ala Val Pro Ala Leu Gly Pro 330 335 340	1181
GGT GTG AAG AAG CGG CGG CAC GGA GAC GAG GAC ACG TAC TAC CTG CAG Gly Val Lys Lys Arg Arg His Gly Asp Glu Asp Thr Tyr Tyr Leu Gln 345 350 355	1229
GTG CGA GGC CGC GAG AAC TTC GAG ATC CTG ATG AAG CTG AAG GAG AGC Val Arg Gly Arg Glu Asn Phe Glu Ile Leu Met Lys Leu Lys Glu Ser 360 365 370	1277
CTG GAG CTG ATG GAG TTG GTG CCG CAG CCG CTG GTA GAC TCC TAT CGG Leu Glu Leu Met Glu Leu Val Pro Gln Pro Leu Val Asp Ser Tyr Arg 375 380 385 390	1325
CAG CAG CAG CAG CTC CTA CAG AGG CCG AGT CAC CTA CAG CCC CCA TCC Gln Gln Gln Leu Leu Gln Arg Pro Ser His Leu Gln Pro Pro Ser 395 400 405	1373
TAC GGG CCG GTC CTC TCG CCC ATG AAC AAG GTG CAC GGG GGC GTG AAC Tyr Gly Pro Val Leu Ser Pro Met Asn Lys Val His Gly Val Asn 410 415 420	1421
AAG CTG CCC TCC GTC AAC CAG CTG GTG GGC CAG CCT CCC CCG CAC AGC	1469

Lys Leu Pro Ser Val Asn Gln Leu Val Gly Gln Pro Pro Pro His Ser		
425	430	435
TCG GCA GCT ACA CCC AAC CTG GGA CCT GTG GGC TCT GGG ATG CTC AAC		1517
Ser Ala Ala Thr Pro Asn Leu Gly Pro Val Gly Ser Gly Met Leu Asn		
440	445	450
AAC CAC GGC CAC GCA GTG CCA GCC AAC AGC GAG ATG ACC AGC AGC CAC		1565
Asn His Gly His Ala Val Pro Ala Asn Ser Glu Met Thr Ser Ser His		
455	460	465
GGC ACC CAG TCC ATG GTC TCG GGG TCC CAC TGC ACT CCG CCA CCC CCC		1613
Gly Thr Gln Ser Met Val Ser Gly Ser His Cys Thr Pro Pro Pro		
475	480	485
TAC CAC GCC GAC CCC AGC CTC GTC AGG ACC TGG GGG CCC TGAAGATCCC		1662
Tyr His Ala Asp Pro Ser Leu Val Arg Thr Trp Gly Pro		
490	495	
CGAGCAGTAT CGCATGACCA TCTGGCGGGG CCTGCAGGAC CTGAAGCAGG GCCACCGACTA		1722
CGGCGCCGCC CGGCAGCAGC TGCTCCGCTC CAGCAACGCG GCCGCCATT CCATCGGCCGG		1782
CTCCGGGGAG CTGCAGCGCC AGCGGGTCAT GGAGGCCGTG CACTTCCGCG TGCGCCACAC		1842
CATCACCAATC CCCAACCGCG CGGGCCCCGG CGCCGGCCCC GACGAGTGGG CGGACTTCGG		1902
CTTCGACCTG CCCGACTGCA AGGCCCGCAA GCAGCCCATC AAGGAGGAGT TCACGGAGGC		1962
CGAGATCCAC TGAGGGGCCG GGCCAGCCA GAGCCTGTGC CACCGCCCAAG AGACCCAGGC		2022
CGCCTCGCTC TC		2034

(2) INFORMATION POUR LA SEQ ID NO: 4:

(i) CARACTERISTIQUES DE LA SEQUENCE:
 (A) LONGUEUR: 499 acides aminés
 (B) TYPE: acide aminé
 (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: protéine

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 4:

Met Ala Gln Ser Thr Thr Ser Pro Asp Gly Gly Thr Thr Phe Glu			
1	5	10	15
His Leu Trp Ser Ser Leu Glu Pro Asp Ser Thr Tyr Phe Asp Leu Pro			
20	25	30	
Gln Ser Ser Arg Gly Asn Asn Glu Val Val Gly Gly Thr Asp Ser Ser			
35	40	45	
Met Asp Val Phe His Leu Glu Gly Met Thr Thr Ser Val Met Ala Gln			
50	55	60	
Phe Asn Leu Leu Ser Ser Thr Met Asp Gln Met Ser Ser Arg Ala Ala			
65	70	75	80
Ser Ala Ser Pro Tyr Thr Pro Glu His Ala Ala Ser Val Pro Thr His			
85	90	95	
Ser Pro Tyr Ala Gln Pro Ser Ser Thr Phe Asp Thr Met Ser Pro Ala			
100	105	110	
Pro Val Ile Pro Ser Asn Thr Asp Tyr Pro Gly Pro His His Phe Glu			
115	120	125	
Val Thr Phe Gln Gln Ser Ser Thr Ala Lys Ser Ala Thr Trp Thr Tyr			
130	135	140	

Ser Pro Leu Leu Lys Lys Leu Tyr Cys Gln Ile Ala Lys Thr Cys Pro
 145 150 155 160
 Ile Gln Ile Lys Val Ser Ala Pro Pro Pro Gly Thr Ala Ile Arg
 165 170 175
 Ala Met Pro Val Tyr Lys Lys Ala Glu His Val Thr Asp Ile Val Lys
 180 185 190
 Arg Cys Pro Asn His Glu Leu Gly Arg Asp Phe Asn Glu Gly Gln Ser
 195 200 205
 Ala Pro Ala Ser His Leu Ile Arg Val Glu Gly Asn Asn Leu Ser Gln
 210 215 220
 Tyr Val Asp Asp Pro Val Thr Gly Arg Gln Ser Val Val Val Pro Tyr
 225 230 235 240
 Glu Pro Pro Gln Val Gly Thr Glu Phe Thr Thr Ile Leu Tyr Asn Phe
 245 250 255
 Met Cys Asn Ser Ser Cys Val Gly Gly Met Asn Arg Arg Pro Ile Leu
 260 265 270
 Ile Ile Ile Thr Leu Glu Thr Arg Asp Gly Gln Val Leu Gly Arg Arg
 275 280 285
 Ser Phe Glu Gly Arg Ile Cys Ala Cys Pro Gly Arg Asp Arg Lys Ala
 290 295 300
 Asp Glu Asp His Tyr Arg Glu Gln Gln Ala Leu Asn Glu Ser Ser Ala
 305 310 315 320
 Lys Asn Gly Ala Ala Ser Lys Arg Ala Phe Lys Gln Ser Pro Pro Ala
 325 330 335
 Val Pro Ala Leu Gly Pro Gly Val Lys Lys Arg Arg His Gly Asp Glu
 340 345 350
 Asp Thr Tyr Tyr Leu Gln Val Arg Gly Arg Glu Asn Phe Glu Ile Leu
 355 360 365
 Met Lys Leu Lys Glu Ser Leu Glu Leu Met Glu Leu Val Pro Gln Pro
 370 375 380
 Leu Val Asp Ser Tyr Arg Gln Gln Gln Leu Leu Gln Arg Pro Ser
 385 390 395 400
 His Leu Gln Pro Pro Ser Tyr Gly Pro Val Leu Ser Pro Met Asn Lys
 405 410 415
 Val His Gly Gly Val Asn Lys Leu Pro Ser Val Asn Gln Leu Val Gly
 420 425 430
 Gln Pro Pro Pro His Ser Ser Ala Ala Thr Pro Asn Leu Gly Pro Val
 435 440 445
 Gly Ser Gly Met Leu Asn Asn His Gly His Ala Val Pro Ala Asn Ser
 450 455 460
 Glu Met Thr Ser Ser His Gly Thr Gln Ser Met Val Ser Gly Ser His
 465 470 475 480
 Cys Thr Pro Pro Pro Tyr His Ala Asp Pro Ser Leu Val Arg Thr
 485 490 495
 Trp Gly Pro

(i) CARACTERISTIQUES DE LA SEQUENCE:
 (A) LONGUEUR: 2156 paires de bases
 (B) TYPE: acide nucléique
 (C) NOMBRE DE BRINS: double
 (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc

(vi) ORIGINE:
 (A) ORGANISME: Homo sapiens

(ix) CARACTERISTIQUE ADDITIONELLE:
 (A) NOM/CLE: CDS
 (B) EMPLACEMENT: 33..1940

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 5:

GCGAGCTGCC CTCGGAGGCC GGCGTGGGGA AG ATG GCC CAG TCC ACC GCC ACC Met Ala Gln Ser Thr Ala Thr	53
1 5	
TCC CCT GAT GGG GGC ACC ACG TTT GAG CAC CTC TGG AGC TCT CTG GAA Ser Pro Asp Gly Gly Thr Thr Phe Glu His Leu Trp Ser Ser Leu Glu	101
10 15 20	
CCA GAC AGC ACC TAC TTC GAC CTT CCC CAG TCA AGC CGG GGG AAT AAT Pro Asp Ser Thr Tyr Phe Asp Leu Pro Gln Ser Ser Arg Gly Asn Asn	149
25 30 35	
GAG GTG GTG GGC GGA ACG GAT TCC AGC ATG GAC GTC TTC CAC CTG GAG Glu Val Val Gly Gly Thr Asp Ser Ser Met Asp Val Phe His Leu Glu	197
40 45 50 55	
GGC ATG ACT ACA TCT GTC ATG GCC CAG TTC AAT CTG CTG AGC AGC ACC Gly Met Thr Thr Ser Val Met Ala Gln Phe Asn Leu Leu Ser Ser Thr	245
60 65 70	
ATG GAC CAG ATG AGC AGC CGC GCG GCC TCG GCC AGC CCC TAC ACC CCA Met Asp Gln Met Ser Ser Arg Ala Ala Ser Ala Ser Pro Tyr Thr Pro	293
75 80 85	
GAG CAC GCC AGC GTG CCC ACC CAC TCG CCC TAC GCA CAA CCC AGC Glu His Ala Ala Ser Val Pro Thr His Ser Pro Tyr Ala Gln Pro Ser	341
90 95 100	
TCC ACC TTC GAC ACC ATG TCG CCG GCG CCT GTC ATC CCC TCC AAC ACC Ser Thr Phe Asp Thr Met Ser Pro Ala Pro Val Ile Pro Ser Asn Thr	389
105 110 115	
GAC TAC CCC GGA CCC CAC CAC TTT GAG GTC ACT TTC CAG CAG TCC AGC Asp Tyr Pro Gly Pro His His Phe Glu Val Thr Phe Gln Gln Ser Ser	437
120 125 130 135	
ACG GCC AAG TCA GCC ACC TGG ACG TAC TCC CCG CTC TTG AAG AAA CTC Thr Ala Lys Ser Ala Thr Trp Thr Tyr Ser Pro Leu Leu Lys Lys Leu	485
140 145 150	
TAC TGC CAG ATC GCC AAG ACA TGC CCC ATC CAG ATC AAG GTG TCC ACC Tyr Cys Gln Ile Ala Lys Thr Cys Pro Ile Gln Ile Lys Val Ser Thr	533
155 160 165	
CCG CCA CCC CCA GGC ACT GCC ATC CGG GCC ATG CCT GTT TAC AAG AAA Pro Pro Pro Gly Thr Ala Ile Arg Ala Met Pro Val Tyr Lys Lys	581
170 175 180	
GGC GAG CAC GTG ACC GAC GTC GTG AAA CGC TGC CCC AAC CAC GAG CTC Ala Glu His Val Thr Asp Val Val Lys Arg Cys Pro Asn His Glu Leu	629
185 190 195	
GGG AGG GAC TTC AAC GAA GGA CAG TCT GCT CCA GCC AGC CAC CTC ATC	677

Gly Arg Asp Phe Asn Glu Gly Gln Ser Ala Pro Ala Ser His Leu Ile 200 205 210 215	
CGC GTG GAA GGC AAT AAT CTC TCG CAG TAT GTG GAT GAC CCT GTC ACC Arg Val Glu Gly Asn Asn Leu Ser Gln Tyr Val Asp Asp Pro Val Thr 220 225 230 235 240 245	725
GGC AGG CAG AGC GTC GTG GTG CCC TAT GAG CCA CCA CAG GTG GGG ACG Gly Arg Gln Ser Val Val Val Pro Tyr Glu Pro Pro Gln Val Gly Thr 235 240 245	773
GAA TTC ACC ACC ATC CTG TAC AAC TTC ATG TGT AAC AGC AGC TGT GTA Glu Phe Thr Thr Ile Leu Tyr Asn Phe Met Cys Asn Ser Ser Cys Val 250 255 260	821
GGG GGC ATG AAC CGG CGG CCC ATC CTC ATC ATC ATC ACC CTG GAG ATG Gly Gly Met Asn Arg Arg Pro Ile Leu Ile Ile Ile Thr Leu Glu Met 265 270 275	869
CGG GAT GGG CAG GTG CTG GGC CGC CGG TCC TTT GAG GGC CGC ATC TGC Arg Asp Gly Gln Val Leu Gly Arg Arg Ser Phe Glu Gly Arg Ile Cys 280 285 290 295	917
GCC TGT CCT GGC CGC GAC CGA AAA GCT GAT GAG GAC CAC TAC CGG GAG Ala Cys Pro Gly Arg Asp Arg Lys Ala Asp Glu Asp His Tyr Arg Glu 300 305 310	965
CAG CAG GCC CTG AAC GAG AGC TCC GCC AAG AAC GGG GCC GCC AGC AAG Gln Gln Ala Leu Asn Glu Ser Ser Ala Lys Asn Gly Ala Ala Ser Lys 315 320 325	1013
CGT GCC TTC AAG CAG AGC CCC CCT GCC GTC CCC GCC CTT GGT GCC GGT Arg Ala Phe Lys Gln Ser Pro Pro Ala Val Pro Ala Leu Gly Ala Gly 330 335 340	1061
GTG AAG AAG CGG CGG CAT GGA GAC GAG GAC ACG TAC TAC CTT CAG GTG Val Lys Lys Arg Arg His Gly Asp Glu Asp Thr Tyr Tyr Leu Gln Val 345 350 355	1109
CGA GGC CGG GAG AAC TTT GAG ATC CTG ATG AAG CTG AAA GAG AGC CTG Arg Gly Arg Glu Asn Phe Glu Ile Leu Met Lys Leu Lys Glu Ser Leu 360 365 370 375	1157
GAG CTG ATG GAG TTG GTG CCG CAG CCA CTG GTG GAC TCC TAT CGG CAG Glu Leu Met Glu Leu Val Pro Gln Pro Leu Val Asp Ser Tyr Arg Gln 380 385 390	1205
CAG CAG CAG CTC CTA CAG AGG CCG AGT CAC CTA CAG CCC CCG TCC TAC Gln Gln Gln Leu Leu Gln Arg Pro Ser His Leu Gln Pro Pro Ser Tyr 395 400 405	1253
GGG CCG GTC CTC TCG CCC ATG AAC AAG GTG CAC GGG GGC ATG AAC AAG Gly Pro Val Leu Ser Pro Met Asn Lys Val His Gly Met Asn Lys 410 415 420	1301
CTG CCC TCC GTC AAC CAG CTG GTG GGC CAG CCT CCC CCG CAC AGT TCG Leu Pro Ser Val Asn Gln Leu Val Gly Gln Pro Pro Pro His Ser Ser 425 430 435	1349
GCA GCT ACA CCC AAC CTG GGG CCC GTG GGC CCC GGG ATG CTC AAC AAC Ala Ala Thr Pro Asn Leu Gly Pro Val Gly Pro Gly Met Leu Asn Asn 440 445 450 455	1397
CAT GGC CAC GCA GTG CCA GCC AAC GGC GAG ATG AGC AGC AGC CAC AGC His Gly His Ala Val Pro Ala Asn Gly Glu Met Ser Ser Ser His Ser 460 465 470	1445
GCC CAG TCC ATG GTC TCG GGG TCC CAC TGC ACT CCG CCA CCC CCC TAC Ala Gln Ser Met Val Ser Gly Ser His Cys Thr Pro Pro Pro Pro Tyr 475 480 485	1493
CAC GCC GAC CCC AGC CTC GTC AGT TTT TTA ACA GGA TTG GGG TGT CCA	1541

(2) INFORMATION POUR LA SEQ ID NO: 6:

- (i) CARACTERISTIQUES DE LA SEQUENCE:

 - (A) LONGUEUR: 636 acides aminés
 - (B) TYPE: acide aminé
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: protéine

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 6:

Met Ala Gln Ser Thr Ala Thr Ser Pro Asp Gly Gly Thr Thr Phe Glu
 1 5 10 15
 His Leu Trp Ser Ser Leu Glu Pro Asp Ser Thr Tyr Phe Asp Leu Pro
 20 25 30
 Gln Ser Ser Arg Gly Asn Asn Glu Val Val Gly Gly Thr Asp Ser Ser
 35 40 45
 Met Asp Val Phe His Leu Glu Gly Met Thr Thr Ser Val Met Ala Gln
 50 55 60
 Phe Asn Leu Leu Ser Ser Thr Met Asp Gln Met Ser Ser Arg Ala Ala
 65 70 75 80

Ser Ala Ser Pro Tyr Thr Pro Glu His Ala Ala Ser Val Pro Thr His
85 90 95

Ser Pro Tyr Ala Gln Pro Ser Ser Thr Phe Asp Thr Met Ser Pro Ala
100 105 110

Pro Val Ile Pro Ser Asn Thr Asp Tyr Pro Gly Pro His His Phe Glu
115 120 125

Val Thr Phe Gln Gln Ser Ser Thr Ala Lys Ser Ala Thr Trp Thr Tyr
130 135 140

Ser Pro Leu Leu Lys Lys Leu Tyr Cys Gln Ile Ala Lys Thr Cys Pro
145 150 155 160

Ile Gln Ile Lys Val Ser Thr Pro Pro Pro Gly Thr Ala Ile Arg
165 170 175

Ala Met Pro Val Tyr Lys Lys Ala Glu His Val Thr Asp Val Val Lys
180 185 190

Arg Cys Pro Asn His Glu Leu Gly Arg Asp Phe Asn Glu Gly Gln Ser
195 200 205

Ala Pro Ala Ser His Leu Ile Arg Val Glu Gly Asn Asn Leu Ser Gln
210 215 220

Tyr Val Asp Asp Pro Val Thr Gly Arg Gln Ser Val Val Val Pro Tyr
225 230 235 240

Glu Pro Pro Gln Val Gly Thr Glu Phe Thr Thr Ile Leu Tyr Asn Phe
245 250 255

Met Cys Asn Ser Ser Cys Val Gly Gly Met Asn Arg Arg Pro Ile Leu
260 265 270

Ile Ile Ile Thr Leu Glu Met Arg Asp Gly Gln Val Leu Gly Arg Arg
275 280 285

Ser Phe Glu Gly Arg Ile Cys Ala Cys Pro Gly Arg Asp Arg Lys Ala
290 295 300

Asp Glu Asp His Tyr Arg Glu Gln Gln Ala Leu Asn Glu Ser Ser Ala
305 310 315 320

Lys Asn Gly Ala Ala Ser Lys Arg Ala Phe Lys Gln Ser Pro Pro Ala
325 330 335

Val Pro Ala Leu Gly Ala Gly Val Lys Lys Arg Arg His Gly Asp Glu
340 345 350

Asp Thr Tyr Tyr Leu Gln Val Arg Gly Arg Glu Asn Phe Glu Ile Leu
355 360 365

Met Lys Leu Lys Glu Ser Leu Glu Leu Met Glu Leu Val Pro Gln Pro
370 375 380

Leu Val Asp Ser Tyr Arg Gln Gln Gln Leu Leu Gln Arg Pro Ser
385 390 395 400

His Leu Gln Pro Pro Ser Tyr Gly Pro Val Leu Ser Pro Met Asn Lys
405 410 415

Val His Gly Gly Met Asn Lys Leu Pro Ser Val Asn Gln Leu Val Gly
420 425 430

Gln Pro Pro Pro His Ser Ser Ala Ala Thr Pro Asn Leu Gly Pro Val
435 440 445

Gly Pro Gly Met Leu Asn Asn His Gly His Ala Val Pro Ala Asn Gly
450 455 460

Glu Met Ser Ser Ser His Ser Ala Gln Ser Met Val Ser Gly Ser His
 465 470 475 480
 Cys Thr Pro Pro Pro Tyr His Ala Asp Pro Ser Leu Val Ser Phe
 485 490 495
 Leu Thr Gly Leu Gly Cys Pro Asn Cys Ile Glu Tyr Phe Thr Ser Gln
 500 505 510
 Gly Leu Gln Ser Ile Tyr His Leu Gln Asn Leu Thr Ile Glu Asp Leu
 515 520 525
 Gly Ala Leu Lys Ile Pro Glu Gln Tyr Arg Met Thr Ile Trp Arg Gly
 530 535 540
 Leu Gln Asp Leu Lys Gln Gly His Asp Tyr Ser Thr Ala Gln Gln Leu
 545 550 555 560
 Leu Arg Ser Ser Asn Ala Ala Thr Ile Ser Ile Gly Gly Ser Gly Glu
 565 570 575
 Leu Gln Arg Gln Arg Val Met Glu Ala Val His Phe Arg Val Arg His
 580 585 590
 Thr Ile Thr Ile Pro Asn Arg Gly Gly Pro Gly Gly Pro Asp Glu
 595 600 605
 Trp Ala Asp Phe Gly Phe Asp Leu Pro Asp Cys Lys Ala Arg Lys Gln
 610 615 620
 Pro Ile Lys Glu Glu Phe Thr Glu Ala Glu Ile His
 625 630 635

(2) INFORMATION POUR LA SEQ ID NO: 7:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (A) LONGUEUR: 2040 paires de bases
 - (B) TYPE: acide nucléique
 - (C) NOMBRE DE BRINS: double
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNC

(vi) ORIGINE:

- (A) ORGANISME: Mus musculus

(ix) CARACTERISTIQUE ADDITIONNELLE:

- (A) NOM/CLE: CDS
- (B) EMPLACEMENT: 124..1890

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 7:

TGATCTCCCT GTGGCCTGCA GGGGACTGAG CCAGGGAGTA GATGCCCTGA GACCCAAGG	60
GACACCCAAG GAAACCTTGC TGGCTTGAG AAAGGGATCG TCTCTCTCCT GCCCAAGAGA	120
AGC ATG TGT ATG GGC CCT GTG TAT GAA TCC TTG GGG CAG GCC CAG TTC	168
Met Cys Met Gly Pro Val Tyr Glu Ser Leu Gly Gin Ala Gln Phe	
1 5 10 15	
AAT TTG CTC AGC AGT GCC ATG GAC CAG ATG GGC AGC CGT GCG GCC CCG	216
Asn Leu Leu Ser Ser Ala Met Asp Gln Met Gly Ser Arg Ala Ala Pro	
20 25 30	
GCG AGC CCC TAC ACC CCG GAG CAC GCC GCC AGC GCG CCC ACC CAC TCG	264
Ala Ser Pro Tyr Thr Pro Glu His Ala Ala Ser Ala Pro Thr His Ser	
35 40 45	
CCC TAC GCG CAG CCC AGC TCC ACC TTC GAC ACC ATG TCT CCG GCG CCT	312
Pro Tyr Ala Gln Pro Ser Ser Thr Phe Asp Thr Met Ser Pro Ala Pro	
50 55 60	

AGT CAC CTG CAG CCT CCA TCC TAT GGG CCC GTG CTC TCC CCA ATG AAC Ser His Leu Gln Pro Pro Ser Tyr Gly Pro Val Leu Ser Pro Met Asn 355 360 365	1224
AAG GTA CAC GGT GTC AAC AAA CTG CCC TCC GTC AAC CAG CTG GTG Lys Val His Gly Gly Val Asn Lys Leu Pro Ser Val Asn Gln Leu Val 370 375 380	1272
GCG CAG CCT CCC CCG CAC AGC TCA GCA GCT GGG CCC AAC CTG GGG CCC Gly Gln Pro Pro Pro His Ser Ser Ala Ala Gly Pro Asn Leu Gly Pro 385 390 395	1320
ATG GGC TCC GGG ATG CTC AAC AGC CAC GGC CAC AGC ATG CCG GCC AAT Met Gly Ser Gly Met Leu Asn Ser His Gly His Ser Met Pro Ala Asn 400 405 410 415	1368
GGT GAG ATG AAT GGA GGC CAC AGC TCC CAG ACC ATG GTT TCG GGA TCC Gly Glu Met Asn Gly Gly His Ser Ser Gln Thr Met Val Ser Gly Ser 420 425 430	1416
CAC TGC ACC CCG CCA CCC CCC TAT CAT GCA GAC CCC AGC CTC GTC AGT His Cys Thr Pro Pro Pro Tyr His Ala Asp Pro Ser Leu Val Ser 435 440 445	1464
TTT TTG ACA GGG TTG GGG TGT CCA AAC TGC ATC GAG TGC TTC ACT TCC Phe Leu Thr Gly Leu Gly Cys Pro Asn Cys Ile Glu Cys Phe Thr Ser 450 455 460	1512
CAA GGG TTG CAG AGC ATC TAC CAC CTG CAG AAC CTT ACC ATC GAG GAC Gln Gly Leu Gln Ser Ile Tyr His Leu Gln Asn Leu Thr Ile Glu Asp 465 470 475	1560
CTT GGG GCT CTG AAG GTC CCT GAC CAG TAC CGT ATG ACC ATC TGG AGG Leu Gly Ala Leu Lys Val Pro Asp Gln Tyr Arg Met Thr Ile Trp Arg 480 485 490 495	1608
GGC CTA CAG GAC CTG AAG CAG AGC CAT GAC TGC GGC CAG CAA CTG CTA Gly Leu Gln Asp Leu Lys Gln Ser His Asp Cys Gly Gln Gln Leu Leu 500 505 510	1656
CGC TCC AGC AGC AAC GCG GCC ACC ATC TCC ATC GGC GGC TCT GGC GAG Arg Ser Ser Ser Asn Ala Ala Thr Ile Ser Ile Gly Gly Ser Gly Glu 515 520 525	1704
CTG CAG CGG CAG CGG GTC ATG GAA GCC GTG CAT TTC CGT GTG CGC CAC Leu Gln Arg Gln Arg Val Met Glu Ala Val His Phe Arg Val Arg His 530 535 540	1752
ACC ATC ACA ATC CCC AAC CGT GGA GGC GCA GGT GCG GTG ACA GGT CCC Thr Ile Thr Ile Pro Asn Arg Gly Gly Ala Gly Ala Val Thr Gly Pro 545 550 555	1800
GAC GAG TGG GCG GAC TTT GGC TTT GAC CTG CCT GAC TGC AAG TCC CGT Asp Glu Trp Ala Asp Phe Gly Phe Asp Leu Pro Asp Cys Lys Ser Arg 560 565 570 575	1848
AAG CAG CCC ATC AAA GAG GAG TTC ACA GAG ACA GAG AGC CAC Lys Gln Pro Ile Lys Glu Glu Phe Thr Glu Thr Glu Ser His 580 585	1890
TGAGGAACGT ACCTTCTTCT CCTGTCCTTC CTCTGTGAGA AACTGCTCTT GGAAAGTGGGA CCTGTTGGCT GTGCCACAG AAACCAGCA GGACCTCTG CGGGATGCCA TTCCCTGAAGG GAAGTCGCTC ATGAACTAAC TCCCTCTTGG	1950 2010 2040

(2) INFORMATION POUR LA SEQ ID NO: 8:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 (A) LONGUEUR: 589 acides aminés

(B) TYPE: acide amine
 (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: protéine

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 8:

Met Cys Met Gly Pro Val Tyr Glu Ser Leu Gly Gln Ala Gln Phe Asn
 1 5 10 15

Leu Leu Ser Ser Ala Met Asp Gln Met Gly Ser Arg Ala Ala Pro Ala
 20 25 30

Ser Pro Tyr Thr Pro Glu His Ala Ala Ser Ala Pro Thr His Ser Pro
 35 40 45

Tyr Ala Gln Pro Ser Ser Thr Phe Asp Thr Met Ser Pro Ala Pro Val
 50 55 60

Ile Pro Ser Asn Thr Asp Tyr Pro Gly Pro His His Phe Glu Val Thr
 65 70 75 80

Phe Gln Gln Ser Ser Thr Ala Lys Ser Ala Thr Trp Thr Tyr Ser Pro
 85 90 95

Leu Leu Lys Lys Leu Tyr Cys Gln Ile Ala Lys Thr Cys Pro Ile Gln
 100 105 110

Ile Lys Val Ser Thr Pro Pro Pro Gly Thr Ala Ile Arg Ala Met
 115 120 125

Pro Val Tyr Lys Lys Ala Glu His Val Thr Asp Ile Val Lys Arg Cys
 130 135 140

Pro Asn His Glu Leu Gly Arg Asp Phe Asn Glu Gly Gln Ser Ala Pro
 145 150 155 160

Ala Ser His Leu Ile Arg Val Glu Gly Asn Asn Leu Ala Gln Tyr Val
 165 170 175

Asp Asp Pro Val Thr Gly Arg Gln Ser Val Val Val Pro Tyr Glu Pro
 180 185 190

Pro Gln Val Gly Thr Glu Phe Thr Thr Ile Leu Tyr Asn Phe Met Cys
 195 200 205

Asn Ser Ser Cys Val Gly Gly Met Asn Arg Arg Pro Ile Leu Val Ile
 210 215 220

Ile Thr Leu Glu Thr Arg Asp Gly Gln Val Leu Gly Arg Arg Ser Phe
 225 230 235 240

Glu Gly Arg Ile Cys Ala Cys Pro Gly Arg Asp Arg Lys Ala Asp Glu
 245 250 255

Asp His Tyr Arg Glu Gln Gln Ala Leu Asn Glu Ser Thr Thr Lys Asn
 260 265 270

Gly Ala Ala Ser Lys Arg Ala Phe Lys Gln Ser Pro Pro Ala Ile Pro
 275 280 285

Ala Leu Gly Thr Asn Val Lys Lys Arg Arg His Gly Asp Glu Asp Met
 290 295 300

Phe Tyr Met His Val Arg Gly Arg Glu Asn Phe Glu Ile Leu Met Lys
 305 310 315 320

Val Lys Glu Ser Leu Glu Leu Met Glu Leu Val Pro Gln Pro Leu Val
 325 330 335

Asp Ser Tyr Arg Gln Gln Gln Gln Gln Leu Leu Gln Arg Pro Ser
 340 345 350

His Leu Gln Pro Pro Ser Tyr Gly Pro Val Leu Ser Pro Met Asn Lys
 355 360 365
 Val His Gly Gly Val Asn Lys Leu Pro Ser Val Asn Gln Leu Val Gly
 370 375 380
 Gln Pro Pro Pro His Ser Ser Ala Ala Gly Pro Asn Leu Gly Pro Met
 385 390 395 400
 Gly Ser Gly Met Leu Asn Ser His Gly His Ser Met Pro Ala Asn Gly
 405 410 415
 Glu Met Asn Gly Gly His Ser Ser Gln Thr Met Val Ser Gly Ser His
 420 425 430
 Cys Thr Pro Pro Pro Pro Tyr His Ala Asp Pro Ser Leu Val Ser Phe
 435 440 445
 Leu Thr Gly Leu Gly Cys Pro Asn Cys Ile Glu Cys Phe Thr Ser Gln
 450 455 460
 Gly Leu Gln Ser Ile Tyr His Leu Gln Asn Leu Thr Ile Glu Asp Leu
 465 470 475 480
 Gly Ala Leu Lys Val Pro Asp Gln Tyr Arg Met Thr Ile Trp Arg Gly
 485 490 495
 Leu Gln Asp Leu Lys Gln Ser His Asp Cys Gly Gln Gln Leu Leu Arg
 500 505 510
 Ser Ser Ser Asn Ala Ala Thr Ile Ser Ile Gly Gly Ser Gly Glu Leu
 515 520 525
 Gln Arg Gln Arg Val Met Glu Ala Val His Phe Arg Val Arg His Thr
 530 535 540
 Ile Thr Ile Pro Asn Arg Gly Ala Gly Ala Val Thr Gly Pro Asp
 545 550 555 560
 Glu Trp Ala Asp Phe Gly Phe Asp Leu Pro Asp Cys Lys Ser Arg Lys
 565 570 575
 Gln Pro Ile Lys Glu Glu Phe Thr Glu Thr Glu Ser His
 580 585

(2) INFORMATION POUR LA SEQ ID NO: 9:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (A) LONGUEUR: 758 paires de bases
 - (B) TYPE: acide nucléique
 - (C) NOMBRE DE BRINS: double
 - (D) CONFIGURATION: linéaire

- (ii) TYPE DE MOLECULE: ADNC

- (vi) ORIGINE:
 - (A) ORGANISME: *Mus musculus*

- (ix) CARACTERISTIQUE ADDITIONNELLE:
 - (A) NOM/CLE: CDS
 - (B) EMPLACEMENT: 389..757

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 9:

TGGTCCCGCT TCGACCAAGA CTCCGGCTAC CAGCTTGCGG GCCCCGCGGA GGAGGAGACC	60
CCGCTGGGGC TAGCTGGCG ACGCCGCAGCA AGCGGCCGCG GGAAGGGAGGC GGGAGGAGCG	120
GGGCCCGAGA CCCCGACTCG GGCAGAGCCA GCTGGGGAGG CGGGGCGCGC GTGGGAGCCA	180

GGGGCCCCGGG	TGGCCGGCCC	TCCTCCGCCA	CGGCTGAGTG	CCCGCGCTGC	CTTCCCCCG	240
GTCCGCCAACG	AAAGGCGCTA	AGCCTGCGGC	AGTCCCCCTCG	CCGCCGCC	CCTGCTCCGC	300
ACCCTTATAA	CCCCCGTCC	CCGATCCAGG	CGAGGAGGCA	ACGGTGCAGC	CCAGCCCTCG	360
CCGACGCCGA	CGCCCCGGCCC	GGAGCAGA	ATG AGC GGC AGC	GTT GGG GAG ATG		412
			Met Ser Gly Ser Val	Gly Glu Met		
			1	5		
GCC CAG ACC TCT TCT TCC TCC TCC ACC TTC GAG CAC CTG TGG AGT						460
Ala Gln Thr Ser Ser Ser Ser Ser Thr Phe Glu His Leu Trp Ser						
10	15	20				
TCT CTA GAG CCA GAC AGC ACC TAC TTT GAC CTC CCC CAG CCC AGC CAA						508
Ser Leu Glu Pro Asp Ser Thr Tyr Phe Asp Leu Pro Gln Pro Ser Gln						
25	30	35				40
GGG ACT AGC GAG GCA TCA GGC AGC GAG GAG TCC AAC ATG GAT GTC TTC						556
Gly Thr Ser Glu Ala Ser Gly Ser Glu Glu Ser Asn Met Asp Val Phe						
45	50	55				
CAC CTG CAA GGC ATG GCC CAG TTC AAT TTG CTC AGC AGT GCC ATG GAC						604
His Leu Gln Gly Met Ala Gln Phe Asn Leu Leu Ser Ser Ala Met Asp						
60	65	70				
CAG ATG GGC AGC CGT GCG GCC CCG GCG AGC CCC TAC ACC CCG GAG CAC						652
Gln Met Gly Ser Arg Ala Ala Pro Ala Ser Pro Tyr Thr Pro Glu His						
75	80	85				
GCC GCC AGC GCG CCC ACC CAC TCG CCC TAC GCG CAG CCC AGC TCC ACC						700
Ala Ala Ser Ala Pro Thr His Ser Pro Tyr Ala Gln Pro Ser Ser Thr						
90	95	100				
TTC GAC ACC ATG TCT CCG GCG CCT GTC ATC CCT TCC AAT ACC GAC TAC						748
Phe Asp Thr Met Ser Pro Ala Pro Val Ile Pro Ser Asn Thr Asp Tyr						
105	110	115				120
CCC GGC CCC C						758
Pro Gly Pro						

(2) INFORMATION POUR LA SEQ ID NO: 10:

- (i) CARACTÉRISTIQUES DE LA SÉQUENCE:

 - (A) LONGUEUR: 123 acides aminés
 - (B) TYPE: acide aminé
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: protéine

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 10:

Met	Ser	Gly	Ser	Val	Gly	Glu	Met	Ala	Gln	Thr	Ser	Ser	Ser	Ser	Ser	
1					5					10						15
Ser	Thr	Phe	Glu	His	Leu	Trp	Ser	Ser	Leu	Glu	Pro	Asp	Ser	Thr	Tyr	
							20			25					30	
Phe	Asp	Leu	Pro	Gln	Pro	Ser	Gln	Gly	Thr	Ser	Glu	Ala	Ser	Gly	Ser	
							35		40					45		
Glu	Glu	Ser	Asn	Met	Asp	Val	Phe	His	Leu	Gln	Gly	Met	Ala	Gln	Phe	
							50		55				60			
Asn	Leu	Leu	Ser	Ser	Ala	Met	Asp	Gln	Met	Gly	Ser	Arg	Ala	Ala	Pro	
							65		70				75			80
Ala	Ser	Pro	Tyr	Thr	Pro	Glu	His	Ala	Ala	Ser	Ala	Pro	Thr	His	Ser	
							85			90					95	

Pro Tyr Ala Gln Pro Ser Ser Thr Phe Asp Thr Met Ser Pro Ala Pro
 100 105 110

Val Ile Pro Ser Asn Thr Asp Tyr Pro Gly Pro
 115 120

(2) INFORMATION POUR LA SEQ ID NO: 11:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (A) LONGUEUR: 559 paires de bases
 - (B) TYPE: acide nucléique
 - (C) NOMBRE DE BRINS: double
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc

(vi) ORIGINE:

(A) ORGANISME: Homo sapiens

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 11:

CGACCTTCCC CAGTCAGGCC	GGGGGAATAA TGAGGTGGTG	GGCGGAACGG ATTCCAGCAT	60
GGACGTCCTTC CACCTGGAGG	GCATGACTAC ATCTGTCATG	CATCCTCGGC TCCTGCCTCA	120
CTAGCTGCGG AGCCTCTCCC	GCTCGGTCCA CGCTGCCGGG	CGGCCACGAC CGTGACCCCTT	180
CCCCCTCGGGC CGCCCGAGTC	CATGCCCTCGT CCCACGGGAC	ACCAGTTCCC TGGCGTGTGC	240
AGACCCCCCG GCGCCTACCA	TGCTGTACGT CGGTGACCCC	GCACGGCACC TCGCCACGGC	300
CCAGTTCAAT CTGCTGAGCA	GCACCATGGA CCAGATGAGC	AGCCGGCGGG CCTCGGCCAG	360
CCCCCTACACC CCAGAGCACG	CCGCCAGCGT GCCCACCCAC	TCGCCCTACG CACAACCCAG	420
CTCCACCTTC GACACCATGT	CGCCGGCGCC TGTCATCCCC	TCCAACACCG ACTACCCGG	480
ACCCCCACAC TTTGAGGTCA	CTTTCAGCA GTCCAGCACG	GCCAAGTCAG CCACCTGGAC	540
GTACTCCCCG CTCTTGAAG			559

(2) INFORMATION POUR LA SEQ ID NO: 12:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (A) LONGUEUR: 1764 paires de bases
 - (B) TYPE: acide nucléique
 - (C) NOMBRE DE BRINS: double
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc

(vi) ORIGINE:

(A) ORGANISME: Homo sapiens

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 12:

ATGCTGTACG TCGGTGACCC	CGCACGGCAC CTCGCCACGG	CCCAGTTCAA TCTGCTGAGC	60
AGCACCATGG ACCAGATGAG	CAGCCGCGCG GCCTGGCCA	GCCCCTACAC CCCAGAGCAC	120
GCCGCCAGCG TGCCCACCCA	CTCGCCCTAC GCACAACCA	GCTCCACCTT CGACACCATG	180
TCGCCGGCGC CTGTCATCCC	CTCCAACACC GACTACCCCG	GACCCACCA CTTTGAGGTC	240
ACTTTCCAGC AGTCCAGCAC	GGCCAAGTCA GCCACCTGGA	CGTACTCCCC GCTCTTGAAG	300
AAACTCTACT GCCAGATCGC	CAAGACATGC CCCATCCAGA	TCAAGGTGTC CACCCCGCCA	360
CCCCCAGGCA CTGCCATCCG	GGCCATGCCT GTTACAAGA	AAGCGGAGCA CGTGACCGAC	420

GTCGTGAAAC GCTGCCCAA CCACGAGCTC	GGGAGGGACT TCAACGAAGG ACAGTCTGCT	480
CCAGCCAGCC ACCTCATCCG CGTGGAAAGGC	AATAATCTCT CGCAGTATGT GGATGACCCT	540
GTCACCGGCA GCCAGAGCGT CGTGGTGCCC	TATGACCCAC CACAGGTGGG GACGGAATT	600
ACCACCATCC TGTACAACCTT CATGTGTAAC	AGCAGCTGTG TAGGGGGCAT GAACCGGGGG	660
CCCATCCTCA TCATCATCAC CCTGGAGATG	CGGGATGGGC AGGTGCTGGG CGCCGGTCC	720
TTTGAGGGCC GCATCTGCGC CTGTCTGGC	CGCGACCGAA AAGCTGATGA GGACCACTAC	780
CGGGAGCAGC AGGCCCTGAA CGAGAGCTCC	GCCAAGAACG GGGCCGCCAG CAAGCGTGCC	840
TTCAAGCAGA GCCCCCTGTC CGTCCCCGCC	CTTGGTGCCG GTGTGAAGAA GCGGCGGCAT	900
GGAGACGAGG ACACGTACTA CCTTCAGGTG	CGAGGCCGGG AGAAACTTTGA GATCCTGATG	960
AAGCTGAAAG AGAGCCTGGA GCTGATGGAG	TTGGTGCCGC AGCCACTGTT GGACTCCTAT	1020
CGGCAGCAGC AGCAGCTCCT ACAGAGGCCG	AGTCACCTAC AGCCCCCGTC CTACGGGCCG	1080
GTCCTCTCGC CCATGAACAA GGTGCACGGG	GGCATGAACA AGCTGCCCTC CGTCAACCAG	1140
CTGGTGGGCC AGCCTCCCCC GCACAGTTCG	GCAGCTACAC CCAACCTGGG GCCCCTGGC	1200
CCCCGGATGC TCAACAACCA TGGCCACGCA	GTGCCAGCCA ACGGCGAGAT GAGCAGCAGC	1260
CACAGGGCCC AGTCCATGGT CTGGGGTCC	CACTGCACTC CGGCCACCCCC CTACCAACGCC	1320
GACCCAGGCC TCGTCAGTTT TTAAACAGGA	TTGGGGTGTCA CAAACTGCAT CGAGTATTC	1380
ACCTCCCAAG GGTTACAGAG CATTACAC	CTGCAGAACCT TGACCATTGA GGACCTGGGG	1440
GCCCTGAAGA TCCCCGAGCA GTACCGCATG	ACCATCTGGC GGGGCCTGCA GGACCTGAAG	1500
CAGGGCCACG ACTACAGCAC CGCGCAGCAG	CTGCTCCGCT CTAGCAACGC GGCCACCATC	1560
TCCATCGGCG GCTCAGGGGA ACTGCAGGCC	CAGCGGGTCA TGGAGGCCGT GCACCTCCGC	1620
GTGCGCCACA CCATCACCAC CCCAACCGC	GGCGGCCAG GCGGCGGCCCG TGACGAGTGG	1680
CGGGACTTCG GCTTCGACCT GCCCCACTGC	AAGGCCGCAG AGCAGCCCAT CAAGGAGGAG	1740
TTCACGGAGG CCGAGATCCA CTGA		1764

(2) INFORMATION POUR LA SEQ ID NO: 13:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (A) LONGUEUR: 587 acides aminés
 - (B) TYPE: acide aminé
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: protéine

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 13:

Met	Leu	Tyr	Val	Gly	Asp	Pro	Ala	Arg	His	Leu	Ala	Thr	Ala	Gln	Phe	
1				5					10					15		
Asn	Leu	Leu	Ser	Ser	Thr	Met	Asp	Gln	Met	Ser	Ser	Arg	Ala	Ala	Ser	
						20		25		30						
Ala	Ser	Pro	Tyr	Thr	Pro	Glu	His	Ala	Ala	Ser	Val	Pro	Thr	His	Ser	
						35		40		45						
Pro	Tyr	Ala	Gln	Pro	S	r	Ser	Thr	Phe	Asp	Thr	Met	Ser	Pro	Ala	Pro
					50		55		60							

Val Ile Pro Ser Asn Thr Asp Tyr Pro Gly Pro His His Phe Glu Val
 65 70 75 80
 Thr Phe Gln Gln Ser Ser Thr Ala Lys Ser Ala Thr Trp Thr Tyr Ser
 85 90 95
 Pro Leu Leu Lys Lys Leu Tyr Cys Gln Ile Ala Lys Thr Cys Pro Ile
 100 105 110
 Gln Ile Lys Val Ser Thr Pro Pro Pro Gly Thr Ala Ile Arg Ala
 115 120 125
 Met Pro Val Tyr Lys Lys Ala Glu His Val Thr Asp Val Val Lys Arg
 130 135 140
 Cys Pro Asn His Glu Leu Gly Arg Asp Phe Asn Glu Gly Gln Ser Ala
 145 150 155 160
 Pro Ala Ser His Leu Ile Arg Val Glu Gly Asn Asn Leu Ser Gln Tyr
 165 170 175
 Val Asp Asp Pro Val Thr Gly Arg Gln Ser Val Val Val Pro Tyr Glu
 180 185 190
 Pro Pro Gln Val Gly Thr Glu Phe Thr Thr Ile Leu Tyr Asn Phe Met
 195 200 205
 Cys Asn Ser Ser Cys Val Gly Gly Met Asn Arg Arg Pro Ile Leu Ile
 210 215 220
 Ile Ile Thr Leu Glu Met Arg Asp Gly Gln Val Leu Gly Arg Arg Ser
 225 230 235 240
 Phe Glu Gly Arg Ile Cys Ala Cys Pro Gly Arg Asp Arg Lys Ala Asp
 245 250 255
 Glu Asp His Tyr Arg Glu Gln Gln Ala Leu Asn Glu Ser Ser Ala Lys
 260 265 270
 Asn Gly Ala Ala Ser Lys Arg Ala Phe Lys Gln Ser Pro Pro Ala Val
 275 280 285
 Pro Ala Leu Gly Ala Gly Val Lys Lys Arg Arg His Gly Asp Glu Asp
 290 295 300
 Thr Tyr Tyr Leu Gln Val Arg Gly Arg Glu Asn Phe Glu Ile Leu Met
 305 310 315 320
 Lys Leu Lys Glu Ser Leu Glu Leu Met Glu Leu Val Pro Gln Pro Leu
 325 330 335
 Val Asp Ser Tyr Arg Gln Gln Gln Leu Leu Gln Arg Pro Ser His
 340 345 350
 Leu Gln Pro Pro Ser Tyr Gly Pro Val Leu Ser Pro Met Asn Lys Val
 355 360 365
 His Gly Gly Met Asn Lys Leu Pro Ser Val Asn Gln Leu Val Gly Gln
 370 375 380
 Pro Pro Pro His Ser Ser Ala Ala Thr Pro Asn Leu Gly Pro Val Gly
 385 390 395 400
 Pro Gly Met Leu Asn Asn His Gly His Ala Val Pro Ala Asn Gly Glu
 405 410 415
 Met Ser Ser Ser His Ser Ala Gln Ser Met Val Ser Gly Ser His Cys
 420 425 430
 Thr Pro Pro Pro Pro Tyr His Ala Asp Pro Ser Leu Val Ser Phe Leu
 435 440 445

Thr Gly Leu Gly Cys Pro Asn Cys Ile Glu Tyr Phe Thr Ser Gln Gly
 450 455 460
 Leu Gln Ser Ile Tyr His Leu Gln Asn Leu Thr Ile Glu Asp Leu Gly
 465 470 475 480
 Ala Leu Lys Ile Pro Glu Gln Tyr Arg Met Thr Ile Trp Arg Gly Leu
 485 490 495
 Gln Asp Leu Lys Gln Gly His Asp Tyr Ser Thr Ala Gln Gln Leu Leu
 500 505 510
 Arg Ser Ser Asn Ala Ala Thr Ile Ser Ile Gly Gly Ser Gly Glu Leu
 515 520 525
 Gln Arg Gln Arg Val Met Glu Ala Val His Phe Arg Val Arg His Thr
 530 535 540
 Ile Thr Ile Pro Asn Arg Gly Gly Pro Gly Gly Pro Asp Glu Trp
 545 550 555 560
 Ala Asp Phe Gly Phe Asp Leu Pro Asp Cys Lys Ala Arg Lys Gln Pro
 565 570 575
 Ile Lys Glu Glu Phe Thr Glu Ala Glu Ile His
 580 585

(2) INFORMATION POUR LA SEQ ID NO: 14:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (A) LONGUEUR: 1521 paires de bases
 - (B) TYPE: acide nucléique
 - (C) NOMBRE DE BRINS: double
 - (D) CONFIGURATION: linéaire
- (ii) TYPE DE MOLECULE: ADNc
- (vi) ORIGINE:
 - (A) ORGANISME: Homo sapiens

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 14:

ATGCTGTACG TCGGTGACCC CGCACGGCAC CTGCCACGG CCCAGTTCAA TCTGCTGAGC	60
AGCACCATGG ACCAGATGAG CAGCCGGCG GCCTGGCCA GCCCCTACAC CCCAGAGCAC	120
GCCGCCAGCG TGCCCACCCA CTCGCCCTAC GCACAACCCA GCTCCACCTT CGACACCATG	180
TCCGCCGGCGC CTGTCATCCCT CTCCAACACC GACTACCCCG GACCCACCA CTTTGAGGTC	240
ACTTTCCAGC AGTCCAGCAC GGCCAAGTCA GCCACCTGGA CGTACTCCCC GCTCTTGAAG	300
AAACTCTACT GCCAGATCCG CAAGACATGC CCCATCCAGA TCAAGGTGTC CACCCCGCCA	360
CCCCCAGGCA CTGCCATCCG GGCCATGCCT GTTACAAGA AAGCGGAGCA CGTGACCGAC	420
GTCGTGAAAC GCTGCCCAA CCACGAGCTC GGGAGGGACT TCAACGAAGG ACAGTCTGCT	480
CCAGCCAGCC ACCTCATCCG CGTGGAAAGGC AATAATCTCT CGCAGTATGT GGATGACCCCT	540
GTCACCGGCA GGCAGAGCGT CGTGGTCCCC TATGAGCCAC CACAGGTGGG GACGGAATT	600
ACCACCATCC TGTACAACCT CATGTGTAAC AGCAGCTGTG TAGGGGGCAT GAACCGGGCG	660
CCCACCTCA TCATCATCAC CCTGGAGATG CGGGATGGGC AGGTGCTGGG CCGCCGGTCC	720
TTTGAGGGCC GCATCTGCGC CTGTCCTGGC CGCGACCGAA AAGCTGATGA GGACCACTAC	780
CGGGAGCAGC AGGCCCTGAA CGAGAGCTCC GCCAAGAACG GGGCCGCCAG CAAGCGTGCC	840
TTCAAGCAGA GCCCCCTGC CGTCCCCGCC CTTGGTGCCG GTGTGAAGAA GCGGCGGCAT	900

GGAGACGAGG ACACGTACTA CCTTCAGGTG CGAGGCCGGG AGAACTTGA GATCCTGATG	960
AAGCTGAAAG AGAGCCTGGA GCTGATGGAG TTGGTGCCGC AGCCACTGGT GGACTCCTAT	1020
CGGCAGCAGC AGCAGCTCCT ACAGAGGCCG CCCCGGGATG CTCAACAAACC ATGGCACGC	1080
AGTGCCAGCC AACGGCGAGA TGAGCAGCAG CCACAGCGCC CAGTCCATGG TCTCGGGGTC	1140
CCACTGCAGCT CGGCCACCCCC CCTACCCACGC CGACCCCCAGC CTCGTAGGA CCTGGGGGCC	1200
CTGAAGATCC CCGAGCAGTA CCGCATGACC ATCTGGCGGG GCCTGCAGGA CCTGAAGCAG	1260
GGCCACGACT ACAGCACCGC GCAGCAGCTG CTCCGCTCTA GCAAAGCGGC CACCATCTCC	1320
ATCGGCGGCT CAGGGAACT GCAGGCCAG CGGGTCATGG AGGCCGTGCA CTTCCGCGTG	1380
CGCCACACCA TCACCATCCC CAACCCGGC GGCCCAGGCG GCGGCCCTGA CGAGTGGGCG	1440
GACTTCGGCT TCGACCTGCC CGACTGCAAG GCCCGCAAGC AGCCCATCAA GGAGGAGTTC	1500
ACGGAGGCCG AGATCCACTG A	1521

(2) INFORMATION POUR LA SEQ ID NO: 15:

(i) CARACTERISTIQUES DE LA SEQUENCE:
 (A) LONGUEUR: 506 acides aminés
 (B) TYPE: acide aminé
 (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: protéine

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 15:

Met Leu Tyr Val Gly Asp Pro Ala Arg His Leu Ala Thr Ala Gln Phe			
1	5	10	15
Asn Leu Leu Ser Ser Thr Met Asp Gln Met Ser Ser Arg Ala Ala Ser			
20	25	30	
Ala Ser Pro Tyr Thr Pro Glu His Ala Ala Ser Val Pro Thr His Ser			
35	40	45	
Pro Tyr Ala Gln Pro Ser Ser Thr Phe Asp Thr Met Ser Pro Ala Pro			
50	55	60	
Val Ile Pro Ser Asn Thr Asp Tyr Pro Gly Pro His His Phe Glu Val			
65	70	75	80
Thr Phe Gln Gln Ser Ser Thr Ala Lys Ser Ala Thr Trp Thr Tyr Ser			
85	90	95	
Pro Leu Leu Lys Lys Leu Tyr Cys Gln Ile Ala Lys Thr Cys Pro Ile			
100	105	110	
Gln Ile Lys Val Ser Thr Pro Pro Pro Gly Thr Ala Ile Arg Ala			
115	120	125	
Met Pro Val Tyr Lys Lys Ala Glu His Val Thr Asp Val Val Lys Arg			
130	135	140	
Cys Pro Asn His Glu Leu Gly Arg Asp Phe Asn Glu Gly Gln Ser Ala			
145	150	155	160
Pro Ala Ser His Leu Ile Arg Val Glu Gly Asn Asn Leu Ser Gln Tyr			
165	170	175	
Val Asp Asp Pro Val Thr Gly Arg Gln Ser Val Val Val Pro Tyr Glu			
180	185	190	

Pro Pro Gln Val Gly Thr Glu Phe Thr Thr Ile Leu Tyr Asn Phe Met
 195 200 205
 Cys Asn Ser Ser Cys Val Gly Gly Met Asn Arg Arg Pro Ile Leu Ile
 210 215 220
 Ile Ile Thr Leu Glu Met Arg Asp Gly Gln Val Leu Gly Arg Arg Ser
 225 230 235 240
 Phe Glu Gly Arg Ile Cys Ala Cys Pro Gly Arg Asp Arg Lys Ala Asp
 245 250 255
 Glu Asp His Tyr Arg Glu Gln Gln Ala Leu Asn Glu Ser Ser Ala Lys
 260 265 270
 Asn Gly Ala Ala Ser Lys Arg Ala Phe Lys Gln Ser Pro Pro Ala Val
 275 280 285
 Pro Ala Leu Gly Ala Gly Val Lys Lys Arg Arg His Gly Asp Glu Asp
 290 295 300
 Thr Tyr Tyr Leu Gln Val Arg Gly Arg Glu Asn Phe Glu Ile Leu Met
 305 310 315 320
 Lys Leu Lys Glu Ser Leu Glu Leu Met Glu Leu Val Pro Gln Pro Leu
 325 330 335
 Val Asp Ser Tyr Arg Gln Gln Gln Leu Leu Gln Arg Pro Pro Arg
 340 345 350
 Asp Ala Gln Gln Pro Trp Pro Arg Ser Ala Ser Gln Arg Arg Asp Glu
 355 360 365
 Gln Gln Pro Gln Arg Pro Val His Gly Leu Gly Val Pro Leu His Ser
 370 375 380
 Ala Thr Pro Leu Pro Arg Arg Pro Gln Pro Arg Gln Asp Leu Gly Ala
 385 390 395 400
 Leu Lys Ile Pro Glu Gln Tyr Arg Met Thr Ile Trp Arg Gly Leu Gln
 405 410 415
 Asp Leu Lys Gln Gly His Asp Tyr Ser Thr Ala Gln Gln Leu Leu Arg
 420 425 430
 Ser Ser Asn Ala Ala Thr Ile Ser Ile Gly Gly Ser Gly Glu Leu Gln
 435 440 445
 Arg Gln Arg Val Met Glu Ala Val His Phe Arg Val Arg His Thr Ile
 450 455 460
 Thr Ile Pro Asn Arg Gly Gly Pro Gly Gly Pro Asp Glu Trp Ala
 465 470 475 480
 Asp Phe Gly Phe Asp Leu Pro Asp Cys Lys Ala Arg Lys Gln Pro Ile
 485 490 495
 Lys Glu Glu Phe Thr Glu Ala Glu Ile His
 500 505

(2) INFORMATION POUR LA SEQ ID NO: 16:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (A) LONGUEUR: 1870 paires de bases
 - (B) TYPE: acide nucléique
 - (C) NOMBRE DE BRINS: double
 - (D) CONFIGURATION: linéaire
- (ii) TYPE DE MOLECULE: ADNc
- (vi) ORIGINE:
 - (A) ORGANISME: Homo sapiens

(ix) CARACTERISTIQUE ADDITIONNELLE:

(A) NOM/CLE: CDS

(B) EMPLACEMENT: 104..1867

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 16:

TGCCCGGGGC TGCGACGGCT GCAGGGAAACC AGACAGCACC TACTTCGACC TTCCCCAGTC	60
AAGCCGGGGG AATAATGAGG TGGTGGGCGG AACGGATTCC AGC ATG GAC GTC TTC Met Asp Val Phe	115 1
CAC CTG GAG GGC ATG ACT ACA TCT GTC ATG GCC CAG TTC AAT CTG CTG His Leu Glu Gly Met Thr Thr Ser Val Met Ala Gln Phe Asn Leu Leu 5 10 15 20	163
AGC AGC ACC ATG GAC CAG ATG AGC AGC CGC GCG GCC TCG GCC AGC CCC Ser Ser Thr Met Asp Gln Met Ser Ser Arg Ala Ala Ser Ala Ser Pro 25 30 35	211
TAC ACC CCA GAG CAC GCC GCC AGC GTG CCC ACC CAC TCG CCC TAC GCA Tyr Thr Pro Glu His Ala Ala Ser Val Pro Thr His Ser Pro Tyr Ala 40 45 50	259
CAA CCC AGC TCC ACC TTC GAC ACC ATG TCG CCG GCG CCT GTC ATC CCC Gln Pro Ser Ser Thr Phe Asp Thr Met Ser Pro Ala Pro Val Ile Pro 55 60 65	307
TCC AAC ACC GAC TAC CCC GGA CCC CAC CAC TTT GAG GTC ACT TTC CAG Ser Asn Thr Asp Tyr Pro Gly Pro His His Phe Glu Val Thr Phe Gln 70 75 80	355
CAG TCC AGC ACG GCC AAG TCA GCC ACC TGG ACG TAC TCC CCG CTC TTG Gln Ser Ser Thr Ala Lys Ser Ala Thr Trp Thr Tyr Ser Pro Leu Leu 85 90 95 100	403
AAG AAA CTC TAC TGC CAG ATC GCC AAG ACA TGC CCC ATC CAG ATC AAG Lys Lys Leu Tyr Cys Gln Ile Ala Lys Thr Cys Pro Ile Gln Ile Lys 105 110 115	451
GTC TCC ACC CCG CCA CCC CCA GGC ACT GCC ATC CGG GCC ATG CCT GTT Val Ser Thr Pro Pro Pro Pro Gly Thr Ala Ile Arg Ala Met Pro Val 120 125 130	499
TAC AAG AAA GCG GAG CAC GTG ACC GAC GTC GTG AAA CGC TGC CCC AAC Tyr Lys Lys Ala Glu His Val Thr Asp Val Val Lys Arg Cys Pro Asn 135 140 145	547
CAC GAG CTC GGG AGG GAC TTC AAC GAA GGA CAG TCT GCT CCA GCC AGC His Glu Leu Gly Arg Asp Phe Asn Glu Gly Gln Ser Ala Pro Ala Ser 150 155 160	595
CAC CTC ATC CGC GTG GAA GGC AAT AAT CTC TCG CAG TAT GTG GAT GAC His Leu Ile Arg Val Glu Gly Asn Asn Leu Ser Gln Tyr Val Asp Asp 165 170 175 180	643
CCT GTC ACC GGC AGG CAG AGC GTC GTG GTG CCC TAT GAG CCA CCA CAG Pro Val Thr Gly Arg Gln Ser Val Val Val Pro Tyr Glu Pro Pro Gln 185 190 195	691
GTG GGG ACG GAA TTC ACC ACC ATC CTG TAC AAC TTC ATG TGT AAC AGC Val Gly Thr Glu Phe Thr Thr Ile Leu Tyr Asn Phe Met Cys Asn Ser 200 205 210	739
AGC TGT GTA GGG GGC ATG AAC CGG CGG CCC ATC CTC ATC ATC ATC ACC Ser Cys Val Gly Gly Met Asn Arg Arg Pro Ile Leu Ile Ile Thr 215 220 225	787
CTG GAG ATG CGG GAT GGG CAG GTG CTG GGC CGC CGG TCC TTT GAG GGC Leu Glu Met Arg Asp Gly Gln Val Leu Gly Arg Arg Ser Phe Glu Gly	835

230	235	240	
CGC ATC TGC GCC TGT CCT GGC CGC GAC CGA AAA GCT GAT GAG GAC CAC Arg Ile Cys Ala Cys Pro Gly Arg Asp Arg Lys Ala Asp Glu Asp His 245 250 255 260			883
TAC CGG GAG CAG CAG GCC CTG AAC GAG AGC TCC GCC AAG AAC GGG GCC Tyr Arg Glu Gln Gln Ala Leu Asn Glu Ser Ser Ala Lys Asn Gly Ala 265 270 275			931
GCC AGC AAG CGT GCC TTC AAG CAG AGC CCC CCT GCC GTC CCC GCC CTT Ala Ser Lys Arg Ala Phe Lys Gln Ser Pro Pro Ala Val Pro Ala Leu 280 285 290			979
GGT GCC GGT GTG AAG AAG CGG CGG CAT GGA GAC GAG GAC ACG TAC TAC Gly Ala Gly Val Lys Lys Arg Arg His Gly Asp Glu Asp Thr Tyr Tyr 295 300 305			1027
CTT CAG GTG CGA GGC CGG GAG AAC TTT GAG ATC CTG ATG AAG CTG AAA Leu Gln Val Arg Gly Arg Glu Asn Phe Glu Ile Leu Met Lys Leu Lys 310 315 320			1075
GAG AGC CTG GAG CTG ATG GAG TTG GTG CCG CAG CCA CTG GTG GAC TCC Glu Ser Leu Glu Leu Met Glu Leu Val Pro Gln Pro Leu Val Asp Ser 325 330 335 340			1123
TAT CGG CAG CAG CAG CAG CTC CTA CAG AGG CCG AGT CAC CTA CAG CCC Tyr Arg Gln Gln Gln Leu Leu Gln Arg Pro Ser His Leu Gln Pro 345 350 355			1171
CCG TCC TAC GGG CCG GTC CTC TCG CCC ATG AAC AAG GTG CAC GGG GGC Pro Ser Tyr Gly Pro Val Leu Ser Pro Met Asn Lys Val His Gly Gly 360 365 370			1219
ATG AAC AAG CTG CCC TCC GTC AAC CAG CTG GTG GGC CAG CCT CCC CCG Met Asn Lys Leu Pro Ser Val Asn Gln Leu Val Gly Gln Pro Pro Pro 375 380 385			1267
CAC AGT TCG GCA GCT ACA CCC AAC CTG GGG CCC GTG GGC CCC GGG ATG His Ser Ser Ala Ala Thr Pro Asn Leu Gly Pro Val Gly Pro Gly Met 390 395 400			1315
CTC AAC AAC CAT GGC CAC GCA GTG CCA GCC AAC GGC GAG ATG AGC AGC Leu Asn Asn His Gly His Ala Val Pro Ala Asn Gly Glu Met Ser Ser 405 410 415 420			1363
AGC CAC AGC GCC CAG TCC ATG GTC TCG GGG TCC CAC TGC ACT CCG CCA Ser His Ser Ala Gln Ser Met Val Ser Gly Ser His Cys Thr Pro Pro 425 430 435			1411
CCC CCC TAC CAC GCC GAC CCC AGC CTC GTC AGT TTT TTA ACA GGA TTG Pro Pro Tyr His Ala Asp Pro Ser Leu Val Ser Phe Leu Thr Gly Leu 440 445 450			1459
GGG TGT CCA AAC TGC ATC GAG TAT TTC ACC TCC CAA GGG TTA CAG AGC Gly Cys Pro Asn Cys Ile Glu Tyr Phe Thr Ser Gln Gly Leu Gln Ser 455 460 465			1507
ATT TAC CAC CTG CAG AAC CTG ACC ATT GAG GAC CTG GGG GCC CTG AAG Ile Tyr His Leu Gln Asn Leu Thr Ile Glu Asp Leu Gly Ala Leu Lys 470 475 480			1555
ATC CCC GAG CAG TAC CGC ATG ACC ATC TGG CGG GGC CTG CAG GAC CTG Ile Pro Glu Gln Tyr Arg Met Thr Ile Trp Arg Gly Leu Gln Asp Leu 485 490 495 500			1603
AAG CAG GGC CAC GAC TAC AGC ACC GCG CAG CAG CTG CTC CGC TCT AGC Lys Gln Gly His Asp Tyr Ser Thr Ala Gln Gln Leu Leu Arg Ser Ser 505 510 515			1651
AAC GCG GCC ACC ATC TCC ATC GGC GGC TCA GGG GAA CTG CAG CGC CAG Asn Ala Ala Thr Ile Ser Ile Gly Gly Ser Gly Glu Leu Gln Arg Gln			1699

520	525	530	
CGG GTC ATG GAG GCC GTG CAC TTC CGC GTG CGC CAC ACC ATC ACC ATC			
Arg Val Met Glu Ala Val His Phe Arg Val Arg His Thr Ile Thr Ile			
535	540	545	1747
CCC AAC CGC GGC GGC CCA GGC GGC CCT GAC GAG TGG GCG GAC TTC			
Pro Asn Arg Gly Gly Pro Gly Gly Pro Asp Glu Trp Ala Asp Phe			
550	555	560	1795
GCG TTC GAC CTG CCC GAC TGC AAG GCC CGC AAG CAG CCC ATC AAG GAG			
Gly Phe Asp Leu Pro Asp Cys Lys Ala Arg Lys Gln Pro Ile Lys Glu			
565	570	575	1843
GAG TTC ACG GAG GCC GAG ATC CAC TGA			
Glu Phe Thr Glu Ala Glu Ile His			
585			1870

(2) INFORMATION POUR LA SEQ ID NO: 17:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 588 acides aminés
- (B) TYPE: acide aminé
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: protéine

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 17:

Met Asp Val Phe His Leu Glu Gly Met Thr Thr Ser Val Met Ala Gln			
1	5	10	15
Phe Asn Leu Leu Ser Ser Thr Met Asp Gln Met Ser Ser Arg Ala Ala			
20	25	30	
Ser Ala Ser Pro Tyr Thr Pro Glu His Ala Ala Ser Val Pro Thr His			
35	40	45	
Ser Pro Tyr Ala Gln Pro Ser Ser Thr Phe Asp Thr Met Ser Pro Ala			
50	55	60	
Pro Val Ile Pro Ser Asn Thr Asp Tyr Pro Gly Pro His His Phe Glu			
65	70	75	80
Val Thr Phe Gln Gln Ser Ser Thr Ala Lys Ser Ala Thr Trp Thr Tyr			
85	90	95	
Ser Pro Leu Leu Lys Lys Leu Tyr Cys Gln Ile Ala Lys Thr Cys Pro			
100	105	110	
Ile Gln Ile Lys Val Ser Thr Pro Pro Pro Gly Thr Ala Ile Arg			
115	120	125	
Ala Met Pro Val Tyr Lys Lys Ala Glu His Val Thr Asp Val Val Lys			
130	135	140	
Arg Cys Pro Asn His Glu Leu Gly Arg Asp Phe Asn Glu Gly Gln Ser			
145	150	155	160
Ala Pro Ala Ser His Leu Ile Arg Val Glu Gly Asn Asn Leu Ser Gln			
165	170	175	
Tyr Val Asp Asp Pro Val Thr Gly Arg Gln Ser Val Val Val Pro Tyr			
180	185	190	
Glu Pro Pro Gln Val Gly Thr Glu Phe Thr Thr Ile Leu Tyr Asn Phe			
195	200	205	
Met Cys Asn Ser Ser Cys Val Gly Met Asn Arg Arg Pro Ile Leu			
210	215	220	

Ile Ile Ile Thr Leu Glu Met Arg Asp Gly Gln Val Leu Gly Arg Arg
 225 230 235 240
 Ser Phe Glu Gly Arg Ile Cys Ala Cys Pro Gly Arg Asp Arg Lys Ala
 245 250 255
 Asp Glu Asp His Tyr Arg Glu Gln Gln Ala Leu Asn Glu Ser Ser Ala
 260 265 270
 Lys Asn Gly Ala Ala Ser Lys Arg Ala Phe Lys Gln Ser Pro Pro Ala
 275 280 285
 Val Pro Ala Leu Gly Ala Gly Val Lys Lys Arg Arg His Gly Asp Glu
 290 295 300
 Asp Thr Tyr Tyr Leu Gln Val Arg Gly Arg Glu Asn Phe Glu Ile Leu
 305 310 315 320
 Met Lys Leu Lys Glu Ser Leu Glu Leu Met Glu Leu Val Pro Gln Pro
 325 330 335
 Leu Val Asp Ser Tyr Arg Gln Gln Gln Leu Leu Gln Arg Pro Ser
 340 345 350
 His Leu Gln Pro Pro Ser Tyr Gly Pro Val Leu Ser Pro Met Asn Lys
 355 360 365
 Val His Gly Gly Met Asn Lys Leu Pro Ser Val Asn Gln Leu Val Gly
 370 375 380
 Gln Pro Pro Pro His Ser Ser Ala Ala Thr Pro Asn Leu Gly Pro Val
 385 390 395 400
 Gly Pro Gly Met Leu Asn Asn His Gly His Ala Val Pro Ala Asn Gly
 405 410 415
 Glu Met Ser Ser Ser His Ser Ala Gln Ser Met Val Ser Gly Ser His
 420 425 430
 Cys Thr Pro Pro Pro Pro Tyr His Ala Asp Pro Ser Leu Val Ser Phe
 435 440 445
 Leu Thr Gly Leu Gly Cys Pro Asn Cys Ile Glu Tyr Phe Thr Ser Gln
 450 455 460
 Gly Leu Gln Ser Ile Tyr His Leu Gln Asn Leu Thr Ile Glu Asp Leu
 465 470 475 480
 Gly Ala Leu Lys Ile Pro Glu Gln Tyr Arg Met Thr Ile Trp Arg Gly
 485 490 495
 Leu Gln Asp Leu Lys Gln Gly His Asp Tyr Ser Thr Ala Gln Gln Leu
 500 505 510
 Leu Arg Ser Ser Asn Ala Ala Thr Ile Ser Ile Gly Gly Ser Gly Glu
 515 520 525
 Leu Gln Arg Gln Arg Val Met Glu Ala Val His Phe Arg Val Arg His
 530 535 540
 Thr Ile Thr Ile Pro Asn Arg Gly Gly Pro Gly Gly Pro Asp Glu
 545 550 555 560
 Trp Ala Asp Phe Gly Phe Asp Leu Pro Asp Cys Lys Ala Arg Lys Gln
 565 570 575
 Pro Ile Lys Glu Glu Phe Thr Glu Ala Glu Ile His
 580 585

(2) INFORMATION POUR LA SEQ ID NO: 18:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 1817 paires de bases
- (B) TYPE: acide nucléique
- (C) NOMBRE DE BRINS: double
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc

(vi) ORIGINE:

- (A) ORGANISME: Homo sapiens

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 18:

ATGGCCCACT CCACCGCCAC CTCCCCGAT GGGGGCACCA CGTTTGGACA CCTCTGGAGC	60
TCTCTGGAAC CAGACAGCAC CTACTTCGAC CTTCCCCAGT CAAGCCGGGG GAATAATGAG	120
GTGGTGGGCC GAACGGATTG CAGCATGGAC GTCTTCCACC TGGAGGGCAT GACTACATCT	180
GTCATGGCCC AGTTCAATCT GCTGAGCAGC ACCATGGACC AGATGAGCAG CGCGCGGGCC	240
TCGGCCAGCC CCTACACCCC AGAGCACGCC GCCAGCGTGC CCACCCACTC GCCCTACGCA	300
CAACCCAGCT CCACCTTCGA CACCATGTGCG CCGGCGCCTG TCATCCCCCTC CAACACCGAC	360
TACCCCGGAC CCCACCACTT TGAGGTCACT TTCCAGCAGT CCAGCACGGC CAAGTCAGCC	420
ACCTGGACGT ACTCCCCGCT CTTGAAGAAA CTCTACTGCC AGATGCCAA GACATGCC	480
ATCCAGATCA AGGTGTCCAC CCCGCCACCC CCAGGCAGTG CCATCCGGGC CATGCC	540
TACAAGAAAG CGGAGCACGT GACCGACGTC GTGAAACGCT GCCCCAACCA CGAGCTCGGG	600
AGGGACTTCA ACGAAGGACA GTCTGCTCCA GCCAGCCACC TCATCCGGT GGAAGGCAAT	660
AATCTCTCGC AGTATGTGGA TGACCCCTGTC ACCGGCAGGC AGAGCGTCGT GGTGCC	720
GAGCCACCAC AGGTGGGAC GGAATTCAAC ACCATCCTGT ACAACTTCAT GTGTAACAGC	780
AGCTGTGTAG GGGGCATGAA CGGGCGGCC ATCCTCATCA TCATCACCCCT GGAGATGCC	840
GATGGGCAGG TGCTGGCCG CGGGTCCCTT GAGGGCCGCA TCTGCGCCTG TCCGGCC	900
GACCGAAAAG CTGATGAGGA CCACTACCGG GAGCAGCAGG CCCTGAACGA GAGCTCC	960
AAGAACGGGG CGGCCAGCAA GCGTGCCTTC AAGCAGAGCC CCCCTGCCGT CCCCCCCTT	1020
GGTGCCGGTG TGAAGAAGCG GCGGCATGGA GACGAGGACA CGTACTACCT TCAGGTGCA	1080
GGCCGGGAGA ACTTGAGAT CCTGATGAACTGAAAGAGA GCCTGGAGCT GATGGAGTTG	1140
GTGCCGCAGC CACTGGTGGCA CTCTATCGG CACGAGCAGC AGCTCCTACA GAGGCCAGT	1200
CACCTACAGC CCCCCGTCTA CGGGCGGCCCTC CTCTGCCCA TGAACAAGGT GCACGGGGC	1260
ATGAACAAGC TGGCCCTCCGT CAACCAGCTG GTGGGCCAGC CTCCCCCGCA CAGTTGGCA	1320
GCTACACCCA ACCTGGGGCC CGTGGGCCCTCA ACAACCATGG CCACGCCAGTG	1380
CCAGCCAACG GCGAGATGAG CAGCAGCCAC AGCGCCAGT CCATGGTCTC GGGGTCCCAC	1440
TGGACTCCGC CACCCCCCTA CCACGCCGAC CCCAGCCTCG TCAGGACCTG GGGGCC	1500
AGATCCCCGA GCAGTACCGC ATGACCATCT GGCGGGCCCT GCAGGACCTG AAGCAGGGC	1560
ACGACTACAG CACCGCGCAG CAGCTGCTCC GCTCTAGCAA CGCGGCCACC ATCTCCATCG	1620
GGGGCTCAGG GGAAGTGCAG CGCCAGCGGG TCATGGAGGC CGTGCAGTTC CGCGTGC	1680
ACACCATCAC CATCCCCAAC CGCGGCCGCG CAGGCCGGCG CCCTGACGAG TGGGCGGACT	1740
TCGGCTTCGA CCTGCCGAC TGCAAGGCC GCAAGCAGCC CATCAAGGAG GAGTTCACGG	1800

AGGCCGAGAT CCACTGA

1817

(2) INFORMATION POUR LA SEQ ID NO: 19:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 (A) LONGUEUR: 499 acides aminés
 (B) TYPE: acide aminé
 (D) CONFIGURATION: linéaire

- (ii) TYPE DE MOLECULE: protéine

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 19:

Met	Ala	Gln	Ser	Thr	Ala	Thr	Ser	Pro	Asp	Gly	Gly	Thr	Thr	Phe	Glu
1										10					15
His Leu Trp Ser Ser Leu Glu Pro Asp Ser Thr Tyr Phe Asp Leu Pro															
										25					30
Gln Ser Ser Arg Gly Asn Asn Glu Val Val Gly Gly Thr Asp Ser Ser															
										40					45
Met Asp Val Phe His Leu Glu Gly Met Thr Thr Ser Val Met Ala Gln															
										55					60
Phe Asn Leu Leu Ser Ser Thr Met Asp Gln Met Ser Ser Arg Ala Ala															
										70					80
Ser Ala Ser Pro Tyr Thr Pro Glu His Ala Ala Ser Val Pro Thr His															
										85					95
Ser Pro Tyr Ala Gln Pro Ser Ser Thr Phe Asp Thr Met Ser Pro Ala															
										100					110
Pro Val Ile Pro Ser Asn Thr Asp Tyr Pro Gly Pro His His Phe Glu															
										115					125
Val Thr Phe Gln Gln Ser Ser Thr Ala Lys Ser Ala Thr Trp Thr Tyr															
										130					140
Ser Pro Leu Leu Lys Lys Leu Tyr Cys Gln Ile Ala Lys Thr Cys Pro															
										145					160
Ile Gln Ile Lys Val Ser Thr Pro Pro Pro Gly Thr Ala Ile Arg															
										165					175
Ala Met Pro Val Tyr Lys Lys Ala Glu His Val Thr Asp Val Val Lys															
										180					190
Arg Cys Pro Asn His Glu Leu Gly Arg Asp Phe Asn Glu Gly Gln Ser															
										195					205
Ala Pro Ala Ser His Leu Ile Arg Val Glu Gly Asn Asn Leu Ser Gln															
										210					220
Tyr Val Asp Asp Pro Val Thr Gly Arg Gln Ser Val Val Val Pro Tyr															
										225					240
Glu Pro Pro Gln Val Gly Thr Glu Phe Thr Thr Ile Leu Tyr Asn Phe															
										245					255
Met Cys Asn Ser Ser Cys Val Gly Gly Met Asn Arg Arg Pro Ile Leu															
										260					270
Ile Ile Ile Thr Leu Glu Met Arg Asp Gly Gln Val Leu Gly Arg Arg															
										275					285
Ser Phe Glu Gly Arg Ile Cys Ala Cys Pro Gly Arg Asp Arg Lys Ala															
										290					300

Asp Glu Asp His Tyr Arg Glu Gln Gln Ala Leu Asn Glu Ser Ser Ala
 305 310 315 320
 Lys Asn Gly Ala Ala Ser Lys Arg Ala Phe Lys Gln Ser Pro Pro Ala
 325 330 335
 Val Pro Ala Leu Gly Ala Gly Val Lys Lys Arg Arg His Gly Asp Glu
 340 345 350
 Asp Thr Tyr Tyr Leu Gln Val Arg Gly Arg Glu Asn Phe Glu Ile Leu
 355 360 365
 Met Lys Leu Lys Glu Ser Leu Glu Leu Met Glu Leu Val Pro Gln Pro
 370 375 380
 Leu Val Asp Ser Tyr Arg Gln Gln Gln Leu Leu Gln Arg Pro Ser
 385 390 395 400
 His Leu Gln Pro Pro Ser Tyr Gly Pro Val Leu Ser Pro Met Asn Lys
 405 410 415
 Val His Gly Gly Met Asn Lys Leu Pro Ser Val Asn Gln Leu Val Gly
 420 425 430
 Gln Pro Pro Pro His Ser Ser Ala Ala Thr Pro Asn Leu Gly Pro Val
 435 440 445
 Gly Pro Gly Met Leu Asn Asn His Gly His Ala Val Pro Ala Asn Gly
 450 455 460
 Glu Met Ser Ser Ser His Ser Ala Gln Ser Met Val Ser Gly Ser His
 465 470 475 480
 Cys Thr Pro Pro Pro Tyr His Ala Asp Pro Ser Leu Val Arg Thr
 485 490 495
 Trp Gly Pro

(2) INFORMATION POUR LA SEQ ID NO: 20:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (A) LONGUEUR: 17 paires de bases
 - (B) TYPE: acide nucléique
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire
- (ii) TYPE DE MOLECULE: ADN
- (iii) HYPOTHETIQUE: NON
- (iii) ANTI-SENS: NON

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 20:

GCGAGCTGCC CTCGGAG

17

(2) INFORMATION POUR LA SEQ ID NO: 21:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (A) LONGUEUR: 19 paires de bases
 - (B) TYPE: acide nucléique
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire
- (ii) TYPE DE MOLECULE: ADN
- (iii) ANTI-SENS: OUI

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 21:

GGTTCTGCAG GTGACTCAG

19

(2) INFORMATION POUR LA SEQ ID NO: 22:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
(A) LONGUEUR: 18 paires de bases
(B) TYPE: acide nucléique
(C) NOMBRE DE BRINS: simple
(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN

(iii) ANTI-SENS: NON

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 22:

GCCATGCCTG TCTACAAG

18

(2) INFORMATION POUR LA SEQ ID NO: 23:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
(A) LONGUEUR: 18 paires de bases
(B) TYPE: acide nucléique
(C) NOMBRE DE BRINS: simple
(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN

(iii) ANTI-SENS: OUI

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 23:

ACCAGCTGGT TGACGGAG

18

(2) INFORMATION POUR LA SEQ ID NO: 24:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
(A) LONGUEUR: 21 paires de bases
(B) TYPE: acide nucléique
(C) NOMBRE DE BRINS: simple
(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN

(iii) ANTI-SENS: NON

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 24:

GTCAACCAGC TGGTGGGCCA G

21

(2) INFORMATION POUR LA SEQ ID NO: 25:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
(A) LONGUEUR: 16 paires de bases
(B) TYPE: acide nucléique
(C) NOMBRE DE BRINS: simple
(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN

(iii) ANTI-SENS: OUI

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 25:

GTGGATCTCG GCCTCC

16

(2) INFORMATION POUR LA SEQ ID NO: 26:

- (1) CARACTERISTIQUES DE LA SEQUENCE:
(A) LONGUEUR: 17 paires de bases
(B) TYPE: acide nucléique
(C) NOMBRE DE BRINS: simple
(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN

(iii) ANTI-SENS: NON

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 26:

AGGCCGGCGT GGGGAAG

17

(2) INFORMATION POUR LA SEQ ID NO: 27:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
(A) LONGUEUR: 19 paires de bases
(B) TYPE: acide nucléique
(C) NOMBRE DE BRINS: simple
(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN

(iii) ANTI-SENS: OUI

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 27:

CTTGGCGATC TGGCAGTAG

19

(2) INFORMATION POUR LA SEQ ID NO: 28:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
(A) LONGUEUR: 17 paires de bases
(B) TYPE: acide nucléique
(C) NOMBRE DE BRINS: simple
(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN

(iii) ANTI-SENS: NON

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 28:

GCGGCCACGA CGGTGAC

17

(2) INFORMATION POUR LA SEQ ID NO: 29:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
(A) LONGUEUR: 18 paires de bases
(B) TYPE: acide nucléique
(C) NOMBRE DE BRINS: simple
(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN

(iii) ANTI-SENS: OUI

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 29:
GGCAGCTTGG GTCTCTGG

18

(2) INFORMATION POUR LA SEQ ID NO: 30:
(1) CARACTERISTIQUES DE LA SEQUENCE:
(A) LONGUEUR: 18 paires de bases
(B) TYPE: acide nucléique
(C) NOMBRE DE BRINS: simple
(D) CONFIGURATION: linéaire
(ii) TYPE DE MOLECULE: ADN
(iii) ANTI-SENS: NON

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 30:
CTGTACGTCG GTGACCCC

18

(2) INFORMATION POUR LA SEQ ID NO: 31:
(i) CARACTERISTIQUES DE LA SEQUENCE:
(A) LONGUEUR: 18 paires de bases
(B) TYPE: acide nucléique
(C) NOMBRE DE BRINS: simple
(D) CONFIGURATION: linéaire
(ii) TYPE DE MOLECULE: ADN
(iii) ANTI-SENS: OUI

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 31:
TCAGTGGATC TCGGCCTC

18

(2) INFORMATION POUR LA SEQ ID NO: 32:
(i) CARACTERISTIQUES DE LA SEQUENCE:
(A) LONGUEUR: 18 paires de bases
(B) TYPE: acide nucléique
(C) NOMBRE DE BRINS: simple
(D) CONFIGURATION: linéaire
(ii) TYPE DE MOLECULE: ADN
(iii) ANTI-SENS: NON

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 32:
AGGGGACGCA GCGAAACC

18

(2) INFORMATION POUR LA SEQ ID NO: 33:
(i) CARACTERISTIQUES DE LA SEQUENCE:
(A) LONGUEUR: 19 paires de bases
(B) TYPE: acide nucléique
(C) NOMBRE DE BRINS: simple
(D) CONFIGURATION: linéaire
(ii) TYPE DE MOLECULE: ADN
(iii) ANTI-SENS: OUI

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 33:
CCATCAGCTC CAGGCTCTC

19

(2) INFORMATION POUR LA SEQ ID NO: 34:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
(A) LONGUEUR: 18 paires de bases
(B) TYPE: acide nucléique
(C) NOMBRE DE BRINS: simple
(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN

(iii) ANTI-SENS: OUI

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 34:
CCAGGACAGG CGCAGATG

18

(2) INFORMATION POUR LA SEQ ID NO: 35:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
(A) LONGUEUR: 19 paires de bases
(B) TYPE: acide nucléique
(C) NOMBRE DE BRINS: simple
(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN

(iii) ANTI-SENS: OUI

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 35:
GATGAGGTGG CTGGCTGG

19

(2) INFORMATION POUR LA SEQ ID NO: 36:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
(A) LONGUEUR: 19 paires de bases
(B) TYPE: acide nucléique
(C) NOMBRE DE BRINS: simple
(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN

(iii) ANTI-SENS: OUI

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 36:
TGGTCAGGTT CTGCAGGTG

19

(2) INFORMATION POUR LA SEQ ID NO: 37:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
(A) LONGUEUR: 18 paires de bases
(B) TYPE: acide nucléique
(C) NOMBRE DE BRINS: simple
(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN

(iii) ANTI-SENS: NON

78

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 37:

CACCTACTCC AGGGATGC

18

(2) INFORMATION POUR LA SEQ ID NO: 38:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (A) LONGUEUR: 21 paires de bases
 - (B) TYPE: acide nucléique
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN

(iii) ANTI-SENS: OUI

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 38:

AGGAAAATAG AAGCGTCAGT C

21

(2) INFORMATION POUR LA SEQ ID NO: 39:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (A) LONGUEUR: 18 paires de bases
 - (B) TYPE: acide nucléique
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN

(iii) ANTI-SENS: NON

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 39:

CAGGCCCACT TGCCTGCC

18

(2) INFORMATION POUR LA SEQ ID NO: 40:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (A) LONGUEUR: 19 paires de bases
 - (B) TYPE: acide nucléique
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN

(iii) ANTI-SENS: OUI

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 40:

CTGTCCCCAA GCTGATGAG

19

REVENDICATIONS

1. Polypeptide purifié, comprenant une séquence d'acides aminés choisie parmi :
 - a) la séquence SEQ ID n° 2 ;
 - b) la séquence SEQ ID n° 4 ;
 - c) la séquence SEQ ID n° 6 ;
 - d) la séquence SEQ ID n° 8 ;
 - e) la séquence SEQ ID n° 10 ;
 - f) la séquence SEQ ID n° 13 ;
 - g) la séquence SEQ ID n° 15 ;
 - h) la séquence SEQ ID n° 17 ;
 - i) la séquence SEQ ID n° 19 ;

et j) toute séquence biologiquement active dérivée de SEQ ID n° 2, SEQ ID n° 4, SEQ ID n° 6, SEQ ID n° 8, SEQ ID n° 10, SEQ ID n° 13, SEQ ID n° 15, SEQ ID n° 17 ou SEQ ID n° 19.
2. Polypeptide selon la revendication 1, caractérisé en ce qu'il comprend la séquence d'acides aminés choisie parmi SEQ ID n° 6, SEQ ID n° 13, SEQ ID n° 15, SEQ ID n° 17 et SEQ ID n° 19.
3. Polypeptide selon la revendication 1, caractérisé en ce qu'il comprend la séquence comprise entre :
 - le résidu 110 et le résidu 310 de SEQ ID n° 2 ou 6 ;
 - le résidu 60 et le résidu 260 de SEQ ID n° 8.
4. Polypeptide selon la revendication 1, caractérisé en ce qu'il résulte d'un épissage alternatif de l'ARN messager du gène correspondant.
5. Polypeptide selon l'une quelconque des revendications précédentes, caractérisé en ce qu'il s'agit d'un polypeptide recombinant produit sous la forme d'une protéine de fusion.
6. Séquence d'acides nucléiques isolée codant pour un polypeptide selon l'une quelconque des revendications précédentes.

7. Séquence d'acides nucléiques isolée selon la revendication 6, caractérisée en ce qu'elle est choisie parmi :
- a) la séquence SEQ ID n° 1 ;
 - b) la séquence SEQ ID n° 3 ;
 - 5 c) la séquence SEQ ID n° 5 ;
 - d) la séquence SEQ ID n° 7 ;
 - e) la séquence SEQ ID n° 9 ;
 - f) la séquence SEQ ID n° 11 ;
 - 10 g) la séquence SEQ ID n° 12 ;
 - h) la séquence SEQ ID n° 14 ;
 - i) la séquence SEQ ID n° 16 ;
 - j) la séquence SEQ ID n° 18 ;
- 15 k) les séquences d'acides nucléiques capables de s'hybrider spécifiquement à la séquence SEQ ID n° 1, SEQ ID n° 3, SEQ ID n° 5, SEQ ID n° 7, SEQ ID n° 9, SEQ ID n° 11, SEQ ID n° 12, SEQ ID n° 14, SEQ ID n° 16 ou SEQ ID n° 18 ou à leurs séquences complémentaires, ou de s'hybrider spécifiquement à leurs séquences proximales ;
- et l) les séquences dérivées des séquences a), b), c), d), e), f), g), h), i), j) ou k) du fait de la dégénérescence du code génétique, de mutation, de délétion, d'insertion, 20 d'un épissage alternatif ou d'une variabilité allélique.
8. Séquence nucléotidique selon la revendication 6, caractérisée en ce qu'il s'agit d'une séquence choisie parmi SEQ ID n° 5, SEQ ID n° 12, SEQ ID n° 14, SEQ ID n° 16 et SEQ ID n° 18 codant respectivement pour le polypeptide de séquences SEQ ID n° 6, SEQ ID n° 13, SEQ ID n° 15, SEQ ID n° 17 et SEQ ID n° 19.
- 25 9. Vecteur de clonage et/ou d'expression contenant une séquence d'acides nucléiques selon l'une quelconque des revendications 6 à 8.
- 30 10. Vecteur selon la revendication 9, caractérisé en ce qu'il s'agit du plasmide pSE1.
11. Cellule hôte transfectée par un vecteur selon la revendication 9 ou 10.
- 35 12. Cellule hôte transfectée selon la revendication 11, caractérisée en ce qu'il s'agit de *E. coli* MC 1061.

13. Sonde nucléotidique ou amorce nucléotidique caractérisée en ce qu'elle s'hybride spécifiquement avec l'une quelconque des séquences selon les revendications 6 à 8 ou leurs séquences complémentaires ou les ARN messagers correspondants ou les gènes correspondants.
5
14. Sonde ou amorce selon la revendication 13, caractérisée en ce qu'elle comporte au moins 16 nucléotides.
15. Sonde ou amorce selon la revendication 13, caractérisée en ce qu'elle comprend l'intégralité de la séquence du gène codant pour l'un des polypeptides de la revendication 1.
10
16. Sonde ou amorce nucléotidique choisie parmi les oligonucléotides suivants ou leurs complémentaires :
15
SEQ ID n° 20 : GCG AGC TGC CCT CGG AG
SEQ ID n° 21 : GGT TCT GCA GGT GAC TCA G
SEQ ID n° 22 : GCC ATG CCT GTC TAC AAG
SEQ ID n° 23 : ACC AGC TGG TTG ACG GAG
SEQ ID n° 24 : GTC AAC CAG CTG GTG GGC CAG
20
SEQ ID n° 25 : GTG GAT CTC GGC CTC C
SEQ ID n° 26 : AGG CCG GCG TGG GGA AG
SEQ ID n° 27 : CTT GGC GAT CTG GCA GTA G
SEQ ID n° 28 : GCG GCC ACG ACC GTG AC
25
SEQ ID n° 29 : GGC AGC TTG GGT CTC TGG
SEQ ID n° 30 : CTG TAC GTC GGT GAC CCC
SEQ ID n° 31 : TCA GTG GAT CTC GGC CTC
SEQ ID n° 32 : AGG GGA CGC AGC GAA ACC
SEQ ID n° 33 : CCA TCA GCT CCA GGC TCT C
30
SEQ ID n° 34 : CCA GGA CAG GCG CAG ATG
SEQ ID n° 35 : GAT GAG GTG GCT GGC TGG A
SEQ ID n° 36 : TGG TCA GGT TCT GCA GGT G
SEQ ID n° 37 : CAC CTA CTC CAG GGA TGC
SEQ ID n° 38 : AGG AAA ATA GAA GCG TCA GTC
35
SEQ ID n° 39 : CAG GCC CAC TTG CCT GCC
et SEQ ID n° 40 : CTG TCC CCA AGC TGA TGA G

17. Utilisation d'une séquence selon l'une quelconque des revendications 6 à 8 pour la fabrication d'amorces oligonucléotidiques pour des réactions de séquençage ou d'amplification spécifique selon la technique de PCR ou toute variante d celle-ci.
18. Couple d'amorces nucléotidiques, caractérisé en ce qu'il comprend les amorces choisies parmi les séquences suivantes :
- a) amorce sens : GCG AGC TGC CCT CGG AG (SEQ ID n° 20)
amorce antisens : GGT TCT GCA GGT GAC TCA G (SEQ ID n° 21)
- b) amorce sens : GCC ATG CCT GTC TAC AAG (SEQ ID n° 22)
amorce antisens : ACC AGC TGG TTG ACG GAG (SEQ ID n° 23)
- c) amorce sens : GTC AAC CAG CTG GTG GGC CAG (SEQ ID n° 24)
amorce antisens : GTG GAT CTC GGC CTC C (SEQ ID n° 25)
- d) amorce sens : AGG CCG GCG TGG GGA AG (SEQ ID n° 26)
amorce antisens : CTT GGC GAT CTG GCA GTA G (SEQ ID n° 27)
- e) amorce sens : GCG GCC ACG ACC GTG A (SEQ ID n° 28)
amorce antisens : GGC AGC TTG GGT CTC TGG (SEQ ID n° 29)
- f) amorce sens : CTG TAC GTC GGT GAC CCC (SEQ ID n° 30)
amorce antisens : TCA GTG GAT CTC GGC CTC (SEQ ID n° 31)
- g) amorce sens : AGG GGA CGC AGC GAA ACC (SEQ ID n° 32)
amorce antisens : GGC AGC TTG GGT CTC TGG (SEQ ID n° 29)
- h) amorce sens : CCCCCCCCCCCCCCN (où N est égal à G, A ou T)
amorce antisens : CCA TCA GCT CCA GGC TCT C (SEQ ID n° 33)
- i) amorce sens : CCCCCCCCCCCCCCN (où N est égal à G, A ou T)
amorce antisens : CCA GGA CAG GCG CAG ATG (SEQ ID n° 34)

j) amorce sens : CCCCCCCCCCCCN (où N est égal à G, A ou T)
amorce antisens : CTT GGC GAT CTG GCA GTA G (SEQ ID n° 27)

5 k) amorce sens : CAC CTA CTC CAG GGA TGC (SEQ ID n° 37)

amorce antisens : AGG AAA ATA GAA GCG TCA GTC (SEQ ID n° 38)

et l) amorce sens : CAG GCC CAC TTG CCT GCC (SEQ ID n° 39)

amorce antisens : CTG TCC CCA AGC TGA TGA G (SEQ ID n° 40).

10

19. Utilisation d'une séquence selon l'une quelconque des revendications 6 à 8, utilisable en thérapie génique.

15

20. Utilisation d'une séquence selon l'une quelconque des revendications 6 à 8, pour la réalisation de sondes ou d'amorces nucléotidiques de diagnostic, ou d'15 séquences antisens utilisables en thérapie génique.

20

21. Utilisation d'amorces nucléotidiques selon l'une quelconque des revendications 6 à 8 pour le séquençage.

25

22. Utilisation d'une sonde ou amorce selon l'une quelconque des revendications 13 à 16, comme outil de diagnostic *in vitro* pour la détection, par des expériences d'hybridation, des séquences d'acides nucléiques codant pour un polypeptid25 selon l'une quelconque des revendications 1 à 4, dans des échantillons biologiques, ou pour la mise en évidence de synthèses aberrantes ou d'anomalies génétiques.

30

23. Méthode de diagnostic *in vitro* pour la détection de synthèses aberrantes ou d'anomalies génétiques au niveau des séquences d'acides nucléiques codant pour un polypeptide selon l'une quelconque des revendications 1 à 4, caractérisée en ce qu'elle comprend :

35

- la mise en contact d'une sonde nucléotidique selon l'une quelconque des revendications 13 à 16 avec un échantillon biologique dans des conditions permettant la formation d'un complexe d'hybridation entre ladite sonde et la susdite séquence nucléotidique, éventuellement après un étape préalable d'amplification de la susdite séquence nucléotidique ;

- la détection du complexe d'hybridation éventuellement formé ;
- éventuellement le séquençage de la séquence nucléotidique formant le complexe d'hybridation avec la sonde de l'invention.

- 5 24. Utilisation d'une séquence d'acides nucléiques selon l'une quelconque des revendications 6 à 8, pour la production d'un polypeptide recombinant selon l'une quelconque des revendications 1 à 5.
- 10 25. Méthode de production d'une protéine recombinante SR-p70, caractérisée en ce que l'on cultive des cellules transfectées selon la revendication 10 ou 11 dans des conditions permettant l'expression d'un polypeptide recombinant d séquence SEQ ID n° 2, SEQ ID n° 4, SEQ ID n° 6, SEQ ID n° 8, SEQ ID n° 10, SEQ ID n° 13, SEQ ID n° 15, SEQ ID n° 17 ou SEQ ID n° 19 ou tout fragment ou dérivé biologiquement actif, et que l'on récupère ledit polypeptide recombinant.
- 15 26. Anticorps mono ou polyclonaux ou leurs fragments, anticorps chimériques ou immunoconjugués, caractérisés en ce qu'ils sont capables de reconnaître spécifiquement un polypeptide selon l'une quelconque des revendications 1 à 4.
- 20 27. Utilisation des anticorps selon la revendication précédente, pour la purification ou la détection d'un polypeptide selon l'une quelconque des revendications 1 à 4 dans un échantillon biologique.
- 25 28. Procédé de diagnostic *in vitro* de pathologies corrélées à une expression ou une accumulation anormale de protéines SR-p70, notamment les phénomènes de cancérisation, à partir d'un prélèvement biologique, caractérisé en ce que l'on met en contact au moins un anticorps selon la revendication 25 avec ledit prélèvement biologique, dans des conditions permettant la formation éventuelle de complexes immunologiques spécifiques entre une protéine SR-p70 et le ou lesdits anticorps et en ce que l'on détecte les complexes immunologiques spécifiques éventuellement formés.
- 30 29. Kit pour le diagnostic *in vitro* d'une expression ou une accumulation anormale de protéines SR-p70 dans un prélèvement biologique et/ou pour la mesure du taux d'expression de celles-ci dans ledit prélèvement comprenant :
- 35

- 5
- au moins un anticorps selon la revendication 25, éventuellement fixé sur un support,
 - des moyens de révélation de la formation de complexes antigènes/anticorps spécifiques entre une protéine SR-p70 et ledit anticorps et/ou des moyens de quantification de ces complexes.
- 10
30. Méthode pour le diagnostic précoce de la formation des tumeurs caractérisée en ce que l'on met en évidence dans un échantillon de sérum prélevé chez un individu des auto-anticorps dirigés contre une protéine SR-p70 selon les étapes consistant à mettre en contact un échantillon de sérum prélevé chez un individu avec un polypeptide de l'invention, éventuellement fixé sur un support, dans des conditions permettant la formation de complexes immunologiques spécifiques entre ledit polypeptide et les auto-anticorps éventuellement présents dans l'échantillon de sérum, et en ce que l'on détecte les complexes immunologiques spécifiques éventuellement formés.
- 15
31. Méthode de détermination d'une variabilité allélique, d'une mutation, d'un délétion, d'une insertion, d'une perte d'hétérozygotie ou d'une anomalie génétique du gène SR-p70 caractérisée en ce qu'elle utilise au moins un séquence nucléotidique selon l'une quelconque des revendications 6 à 8.
- 20
32. Méthode de détermination d'une variabilité allélique du gène SR-p70 au niveau de la position -30 et -20 par rapport à l'ATG d'initiation de l'exon 2 pouvant être impliquée dans des pathologies et caractérisée en ce qu'elle comprend au moins:
- 25
- une étape au cours de laquelle on procède à l'amplification par PCR de l'exon 2 du gène SR-p70 portant la séquence cible à l'aide de couple d'amorces oligonucléotidiques selon l'une quelconque des revendications 6 à 8;
 - une étape au cours de laquelle on procède au traitement des produits amplifiés par un enzyme de restriction dont le site de coupure correspond à l'allèle recherché;
 - une étape au cours de laquelle on procède à la détection ou au dosage d'au moins l'un des produits de la réaction enzymatique.
- 30
33. Composition pharmaceutique comprenant, à titre de principe actif, un polypeptide selon l'une quelconque des revendications 1 à 4.
- 35

34. Composition pharmaceutique selon la revendication précédente, caractérisée en ce qu'elle comprend un polypeptide selon la revendication 2.

5 35. Composition pharmaceutique contenant un inhibiteur ou un activateur de l'activité du SR-p70.

36. Composition pharmaceutique contenant un polypeptide dérivé d'un polypeptid selon l'une quelconque des revendications 1 à 5 caractérisé en ce qu'il est un inhibiteur ou un activateur du SR-p70.

10

15

20

25

30

35

1 / 36

```

1 : TGCCTCCCCGCCCCGGCACCCGCCCGAGGCCCTGTGCTCTGCGAAGGGG 50
     ||| |||||
1 : .....GGGGCTCCGGGG 12
51 ACGCAGCGAACGCCGGGGCCCGGCCAGGCCGGGACGGACGCCGATG 100
     ||| |||||
13 ACACCTGGCGTCCGGGCTGGAAGCGTGCCTTCAGACGGTGACACGGTT 62
101 CCCGGAGCTGCGACGGCTGCAAGCGAGCTGCCCTCGAGGGCCGGTGTGA 150
     ||| |||||
63 CCCTGAGGATTGGCAGCCAGACTGCTTACGGGTAC...TGCCATGGAGG 109
151 CGAAGATGGCCCAGTCCACCAACCTCCCCGATGGGGCACACGTTT 200
     ||| |||||
110 AGCCGCAGTCAGATCCAGCATCGAGCCCCCTCTGAGTCAGGAAACATTT 159
201 GAGCACCTCTGGAGCTCTCTGGAACAGACAGCACCTACTTCGACCTTC 250
     ||| |||||
160 TCAGACTATGAAAACACTACTTCCTGAAAACAAC.GTTCTGTCCTCTTGC 208
251 CCAGTCAGCCGGGAAATAATGAGGTGGTGGGTGGCACGGATTCCAGCA 300
     ||| |||||
209 CGTCCAAGCGGTGGATGTTGATGCTCTCCGGATGATCTTGACCAA 258
301 TGGACGTCTTCCACCTAGAGGGCATGACCACATCTGTCATGCCAGTC 350
     ||| |||||
259 TGG.....TTAACTGAAGACCCAGGTC 280
351 AATTTGCTGAGCAGCACCATGGACCAGATGAGCAGCCGCGCTGCCPCGGC 400
     ||| |||||
281 CAGATGAAGCTC.....CCAGAATGTCAGAGGCTGCTCCCCACA 319
401 CAGCCCCGTACACCCCGAGCACCCGCCAGCTGCCACCCATTCCACCC 450
     ||| |||||
320 TGGCCCCCACACCAGCAGCTCCATACCCGGCGCCCTGCACCCAGCCC. 368
451 ACGCACAGCCAGCTCCACCTTCGACACCATGTCGCCCGCGCCTGTCATC 500
     ||| |||||
369 .....CTCCCTGGCCCTGTCATCCCTGTC 393
501 CCCTCCAAACACCGACTATCCCGAACCCACCACTTCGAGGTCACTTTCCA 550
     ||| |||||
394 CCTTCCCAGAAAACCTTACCAACGGCAGCTACGGTTCCGTCGGCTTCT 443
551 GCAGTCAGCACGGCCAAGTCAGCCACCTGGACGTACTCCCCACTCTGA 600
     ||| |||||
444 GCATTCTGAAACAGCCAAGTCGTGACTTGCACGTACTCCCTGACCTCA 493
601 AGAAACTCTACTGCCAGATGCCAAGACATGCCCATCCAGATCAAGGTG 650
     ||| |||||
494 ACAAGATGTTTGGCAGCTGCCAAGACCTGCCCTGCGAGCTGTGGTT 543
651 TCCGCCCCACCGCCCCCGGCCAGGCCATCCGGCCATGCCCTGTCACAA 700
     ||| |||||
544 GATTCACACCCCCGGCCCGCAGCGCGTCCGGCCATGCCATCTACAA 593
701 GAAGGGGGAGCACGTGACCGACATCGTAAGGCCCTGCCCAACCCAGGAC 750
     ||| |||||
594 GCAGTCACGACATGACTGAGGTGTGAGGGCTGCCCTCCACCATGAGC 643
751 TCGGGAGGGACTTCACCGAAGGACAGTCGCCAGCCAGCCACCTCATC 800
     ||| |||||
644 GCTGCTCAGACAGCGATGGA.....CTGGCCCTCTCAACATCTTATC 687
801 CGTGTGGAAGGCAATAATCTCTCGCAGTATGTGGACGACCCCTGTACCCG 850
     ||| |||||
688 CGAGTGGAAGGAAATTGGCTGTGGAGTATTCGGATGACAGAAAACATTT 737
851 CAGGCACAGCGTCGTGGTGCCTATGAGCCACACAGGTGGGACAGAAT 900
     ||| |||||
738 TCGACATAGTGTGGTGGTGCCTATGAGCCGCTGAGGTGGCTGTACT 787

```

FIG.1

901 TCACCAACCATCCTGTACAACCTCATGTGTAACAGCAGCTGTGTGGGGGGC 950
 788 GTACCAACCATCCACTACAACATGTGTAACAGTTCCTGCATGGCGGC 837
 951 ATGAACCGACGGCCCACATCCTCATCATCACACCCCTGGAGACGGGGATGG 1000
 838 ATGAACCGGAGGCCACATCCTCACAAATTACACTGGAAAGACTCCAGTGG 887
 1001 GCAGGTGCTGGGCCCGGTCCTCGAGGGCCGCATCTGCCCTGTCCCTG 1050
 888 TAATCTACTGGGACGGAACAGCTTGAGGTGGAGTTTGCCCTGTCCCTG 937
 1051 GCGCGACCGAAAAGCCGATGAGGACCACTACCGGGAGCAGCAGGCCCTTG 1100
 938 GGAGAGACCGGCGCACAGAGGAAGAGAATTTC.....G 971
 1101 AATGAGAGCTCGCCAAGAACGGGCTGCCAGCAAGCCGCCTTCAGCA 1150
 972 CAAGAAAGGGAGCCCTGCCACGAGCTGCCCTGGGAGCACTAACGGAG 1021
 1151 GAGTCCCCCTGCCGTCCCCGCCCTGGGCC.GGGTGTGAAGAAGCGGGCG 1199
 1022 CACTGCCAACAACACCAGCTCCTCTCCCAGCCAAGAAAGAAACCACTG 1071
 1200 CACGGAGACGGAGGACACGTACTACCTGCAGGTGCGAGGCCGGAGAACCT 1249
 1072 GATGGAGAATATTCAAC.....CTTCAGATCCGGCGGTGAGGGCTT 1115
 1250 CGAGATCCTGATGAAGCTGAAGGGAGGCCCTGGAGCTGATGGAGTTGGTGC 1299
 1116 CGAGATGTTCCGAGAGCTGAATGAGGCCCTGGAACTCAAGGA..... 1157
 1300 CGCAGCCGCTGGTAGACTCCATGGCAGCAGCAGCAGCTCCATACAGAGG 1349
 1158 TGCCCAGGCTGGAAAGAGCCAGCGG..GGAGCAGGGCTCACTCCAGCCA 1205
 1350 CCGAGTCACCTACAGCCCCATCTACGGGCGGCTCTCTGCCCATGAA 1399
 1206 CCTGAAGTCCAAGAAGGGCAATCTACCTCCGCCATAAAAATTCACTGT 1255
 1400 CAAGGTGACGGGGGCGTGAACAAGCTGCCCTCCGTCAACCAAGCTGGTGG 1449
 1256 TCAAGACAGAGGGGCTGACTCAGACTGACATTTC.....TCAGCTTCTTG 1300
 1450 GCCAGCTCCCCCGCACAGCTCGGAGCTACACCCAACCTGGGACCTGTG 1499
 1301 TTCCCCCACTGAGCCCTCCACCCCCATCT.CTCCCTCCCCCTGCCATTGG 1349
 1500 GGCTCTGGGATGCTCAACAACCACGGGCAACGCAAGTGCCAGCCAACAGGGA 1549
 1350 AGTTCTGGGTCTTAAACCCCTTGCTTCAATAGGTGTGTCAAGCAA 1399
 1550 GATGACCAGCAGCCACGGCACCCAGTCCATGGTCTGGGGTCCCACGTCA 1599
 1400 A..... 1400

FIG.1 cont.

3 / 36 .

1 MAQSTTTSPDGTTFEHLWSSLEPDSTYFDLHQSSRCNNEVGGTDSSMD 50
1 .. MEEPQSDPSIEPPLS...QETFSDLWKLLEPNNVLSPLPSQAVD 41

51 VFHLEGMTTSVMAQFNLLSSTMQMSRAASASPYTPEHAASVPTHSPYA 100
42 DLML...SPDDLAQWLTEDPGPDEAPRMSAAPHMAPTPAAPTPA.APAP 87

101 QPSSTFDTMSPAPVIPSNTDYPGPHHFETFQQSSTAKSATWTYSPLLKK 150
88 APSWPL....SSSVPSQKTYHGSYGFRLGLHSGTAKSVTCTYSPDLNK 132

151 LYCQIAKTCPIQIKVSAPPPTAIRAMPVYKKAEHVTDIVKRCPNHELG 200
133 MFCQLAKTCPVQLWVDSTPPPGSRVRAMAIYKQSQHMTEVRRCPHME.. 180

201 RDFNEGQSAPASHLIRVEGNNLSQLYVDDDPVTGRQSVVVYPYEPQVGTEFT 250
181 RCSDSDGLAPPQHLLRVEGNLRVEYSDDRNTFRHSVVVPYEPPEVGSDCT 230

251 TILYNFMCNSSCVGGMNRRPILIIITLETRDGQVLGRRSFEGRICACPGR 300
231 TIHYNYMCNSCMGGMNRRPILTIITLEDSSGNLLCRNSFEVRVCACPGR 280

301 DRKADEDHYREQNALNESSAKNGAASKRAFKQSPPAVPALGPGVKRRHG 350
281 DRRTEEEENFRKKG..EPCHELPPGSKRALPNNTSSSPQ....PKKKPL 323

351 DEDTYYLQVRGRENFIELMKLKESLELMELVPQPLVDSYRQQQQLLQRPS 400
324 DGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEPAGSRAHSSHLLKSKK 373

401 HLQPPSYGPVLSPMNKVHGGVNKLPSVNQLVGQPPPSSAATPNLGPVGS 450
374 GQSTSRRHKKFMPKTEGPDSD..... 393

FIG. 2

4 / 36

```

1 TGCCTCCCCGCCGCGCACCCGCCGGAGGCCGTGCTCTGCGAAGGG 50
1 TGCCTCCCCGCCGCGCACCCGCCGGAGGCCGTGCTCTGCGAAGGG 50
51 ACGCAGCGAAGCCGGGCCGCCAGGCCGGGGACGGACGCCATG 100
51 ACGCAGCGAAGCCGGGCCGCCAGGCCGGGGACGGACGCCATG 100
101 CCCGGAGCTGCGACGGCTGCAGAGCGAGCTGCCCTCGAGGCCGGTGA 150
101 CCCGGAGCTGCGACGGCTGCAGAGCGAGCTGCCCTCGAGGCCGGTGA 150
151 GGAAGATGGCCCAGTCACCACCTCCCCGATGGGGCACCGTT 200
151 GGAAGATGGCCCAGTCACCACCTCCCCGATGGGGCACCGTT 200
201 GAGCACCTCTGGAGCTCTGGAAACCAGACAGCACCTACTTCGACCTTC 250
201 GAGCACCTCTGGAGCTCTGGAAACCAGACAGCACCTACTTCGACCTTC 250
251 CCAGTCAGCCGGGGATAATGAGGTGGTGGCACGGATTCCAGCA 300
251 CCAGTCAGCCGGGGATAATGAGGTGGTGGCACGGATTCCAGCA 300
301 TGGACGTCTTCCACCTAGAGGGCATGACCACATCTGTCATGGCCCAGTTC 350
301 TGGACGTCTTCCACCTAGAGGGCATGACCACATCTGTCATGGCCCAGTTC 350
351 AATTGCTGAGCAGCACCATGGACCAGATGAGCAGCCGCTGCCCTGGC 400
351 AATTGCTGAGCAGCACCATGGACCAGATGAGCAGCCGCTGCCCTGGC 400
401 CAGCCCGTACACCCCGGAGCACGCCGCCAGCGTGCACCATTCAACCT 450
401 CAGCCCGTACACCCCGGAGCACGCCGCCAGCGTGCACCATTCAACCT 450
451 ACGCACAGCCAGCTCCACCTTCGACACCATGTCGCCGCCGCTGTCATC 500
451 ACGCACAGCCAGCTCCACCTTCGACACCATGTCGCCGCCGCTGTCATC 500
501 CCCTCCAAACACCGACTATCCGGACCCCAACCTTCGAGGTCACTTCCA 550
501 CCCTCCAAACACCGACTATCCGGACCCCAACCTTCGAGGTCACTTCCA 550
551 GCAGTCCAGCACGGCAAGTCAGCCACCTGGACGTACTCCCCACTCTGA 600
551 GCAGTCCAGCACGGCAAGTCAGCCACCTGGACGTACTCCCCACTCTGA 600
601 AGAAACTCTACTGCGAGATGCCAAGACATGCCCATCCAGATCAAGGTG 650
601 AGAAACTCTACTGCGAGATGCCAAGACATGCCCATCCAGATCAAGGTG 650
651 TCCGCCCCACCGCCCCCGGGCACCGCATCCGGCCATGCCCTGTACAA 700
651 TCCGCCCCACCGCCCCCGGGCACCGCATCCGGCCATGCCCTGTACAA 700
701 GAAGGGGGAGCACGTGACCGACATCGTGAAAGCGCTGCCAACACGAGC 750
701 GAAGGGGGAGCACGTGACCGACATCGTGAAAGCGCTGCCAACACGAGC 750
751 TCGGGAGGGACTTCAACGAAGGACAGTCGCCAGCCACCTCATC 800
751 TCGGGAGGGACTTCAACGAAGGACAGTCGCCAGCCACCTCATC 800
801 CGTGTGGAAGGCAATAATCTCTGGCACTATGTGGACGACCCGTCAACGG 850
801 CGTGTGGAAGGCAATAATCTCTGGCACTATGTGGACGACCCGTCAACGG 850
851 CAGGCAGAGCGTGTGGTGCCCTATGAGCCACACAGGTGGGACAGAAT 900
851 CAGGCAGAGCGTGTGGTGCCCTATGAGCCACACAGGTGGGACAGAAT 900

```

FIG.3
cont.

- 901 TCACCACCATCCTGTACAACCTCATGTGTAACAGCAGCTGTGTGGGGGC 950
 - |||||
 - 901 TCACCACCATCCTGTACAACCTCATGTGTAACAGCAGCTGTGTGGGGGC 950
 - |||||
 - 951 ATGAACCGACGGCCCATCCTCATCATCACCCCTGGAGACGCCGGATGG 1000
 - |||||
 - 951 ATGAACCGACGGCCCATCCTCATCATCACCCCTGGAGACGCCGGATGG 1000
 - |||||
 - 1001 GCAGGTGCTGGGCCGCCGGTCTCGAGGGCCGCATCTGCGCTGTCTG 1050
 - |||||
 - 1001 GCAGGTGCTGGGCCGCCGGTCTCGAGGGCCGCATCTGCGCTGTCTG 1050
 - |||||
 - 1051 GCGCGACCGAAAAGCCGATGAGGACCAACTACCGGGAGCAGCAGGCCCTG 1100
 - |||||
 - 1051 GCGCGACCGAAAAGCCGATGAGGACCAACTACCGGGAGCAGCAGGCCCTG 1100
 - |||||
 - 1101 AATGAGAGCTCCGCCAAGAACGGGCTGCCAGCAAGCGCGCTTCAGCA 1150
 - |||||
 - 1101 AATGAGAGCTCCGCCAAGAACGGGCTGCCAGCAAGCGCGCTTCAGCA 1150
 - |||||
 - 1151 GAGTCCTCTGCCGTCCCCGCCCTGGGCCGGGTGTGAAGAACGGGGGC 1200
 - |||||
 - 1151 GAGTCCTCTGCCGTCCCCGCCCTGGGCCGGGTGTGAAGAACGGGGGC 1200
 - |||||
 - 1201 ACGGAGACGAGGACACGTACTACCTGCAGGTGCGAGGCCGGAGAACTTC 1250
 - |||||
 - 1201 ACGGAGACGAGGACACGTACTACCTGCAGGTGCGAGGCCGGAGAACTTC 1250
 - |||||
 - 1251 GAGATCCTGATGAAGCTGAAGGAGAGCCTGGAGCTGATGGAGTTGGTGCC 1300
 - |||||
 - 1251 GAGATCCTGATGAAGCTGAAGGAGAGCCTGGAGCTGATGGAGTTGGTGCC 1300
 - |||||
 - 1301 GCAGCCGCTGGTAGACTCCTATCGCAGCAGCAGCTCTACAGAGGC 1350
 - |||||
 - 1301 GCAGCCGCTGGTAGACTCCTATCGCAGCAGCAGCTCTACAGAGGC 1350
 - |||||
 - 1351 CGAGTCACCTACAGCCCCATCCTACGGGCCGGTCTCTGCCCATGAAC 1400
 - |||||
 - 1351 CGAGTCACCTACAGCCCCATCCTACGGGCCGGTCTCTGCCCATGAAC 1400
 - |||||
 - 1401 AAGGTGACGGGGCGTGAACAAGCTGCCCTCCGTCAACCAAGCTGGTGG 1450
 - |||||
 - 1401 AAGGTGACGGGGCGTGAACAAGCTGCCCTCCGTCAACCAAGCTGGTGG 1450
 - |||||
 - 1451 CCAGCCTCCCCGACAGCTCGCAGCTACACCCAACCTGGGACCTGTGG 1500
 - |||||
 - 1451 CCAGCCTCCCCGACAGCTCGCAGCTACACCCAACCTGGGACCTGTGG 1500
 - |||||
 - 1501 GCTCTGGATGCTAACAAACCGGCCACGGCAGTGCAGCCAACAGCGAG 1550
 - |||||
 - 1501 GCTCTGGATGCTAACAAACCGGCCACGGCAGTGCAGCCAACAGCGAG 1550
 - |||||
 - 1551 ATGACCAAGCAGCCACGGCACCCAGTCCATGCTCTGGGGTCCCCACTGCAC 1600
 - |||||
 - 1551 ATGACCAAGCAGCCACGGCACCCAGTCCATGCTCTGGGGTCCCCACTGCAC 1600
 - |||||
 - 1601 TCCGCCACCCCCCTACCAACGCCGACCCCAGCCCTCGTCAGTTTTAACAG 1650
 - |||||
 - 1601 TCCGCCACCCCCCTACCAACGCCGACCCCAGCCCTCGTCAGTTTTAACAG 1637
 -
 -
 - 1701 AGCATTACCAACCTGCAGAACCTGACCATCGAGGACCTGGGGCCCTGAA 1750
 - |||||
 - 1638AGGACCTGGGGCCCTGAA 1656
 - |||||
 - 1751 GATCCCCGAGCAGTATCGCATGACCATCTGGGGGGCCCTGAGGACCTGA 1800
 - |||||

FIG.3
cont.

6 / 36

1657 GATCCCCGAGCAGTATGGCATGACCATCTGGGGGGCTGCAGGACCTGA 1706
-
- 1801 AGCAGGGCCACGACTACGGCGCCGCCGCAGCAGCTGCTCCGCTCCAGC 1850
-
- 1707 AGCAGGGCCACGACTACGGCGCCGCCGCAGCAGCTGCTCCGCTCCAGC 1756
-
- 1851 AACGGCGCCATTCCATCGGCGCTCCGGGGAGCTGCAGGCCAGCG 1900
-
- 1757 AACGGCGCCATTCCATCGGCGCTCCGGGGAGCTGCAGGCCAGCG 1806
-
- 1901 GGTCAATGGAGGCCGTGCACTTCGGGTGCGCCACACCATCACCATCCCCA 1950
-
- 1807 GGTCAATGGAGGCCGTGCACTTCGGGTGCGCCACACCATCACCATCCCCA 1856
-
- 1951 ACCGGCGCCGCCCGGGCGCCGGGGAGCTGCAGGAGCTTC 2000
-
- 1857 ACCGGCGCCGCCCGGGCGCCGGGGAGCTGCAGGAGCTTC 1906
-
- 2001 GACCTGCCCAGTGCAAGGCCGCAAGCAGCCATCAAGGAGGAGTTAAC 2050
-
- 1907 GACCTGCCCAGTGCAAGGCCGCAAGCAGCCATCAAGGAGGAGTTAAC 1956
-
- 2051 GGAGGCCGAGATCCACTGAGGGGCCGGGCCAGCCAGGCCCTGTGCCACC 2100
-
- 1957 GGAGGCCGAGATCCACTGAGGGGCCGGGCCAGCCAGGCCCTGTGCCACC 2006
-
- 2101 GCCCAGAGACCCAGGCCCTCGCTCTC 2128
-
- 2007 GCCCAGAGACCCAGGCCCTCGCTCTC 2034

FIG.3 cont.

7 / 36

-	1	TGCTCCCCGCCGCACCCGCCCGAGGCTGTGCTCTGCGAAGGGACCGAGCGAA	60
-	61	GCGGGGCCCGCGCAGGCCGGCCGGACGGACGCCGATGCCGGAGCTGCGACGGCTGC	120
-	121	AGAGCGAGCTGCCCTCGGAGGCCGGTGTGAGGAAGATGGCCAGTCCACCACCTCCC	180
-	-10	MAQSTTTSP	9
-	181	CCGATGGGGCACCACTTGTAGCACCTCTGGAGCTCTCTGAAACCAGACAGCACCTACT	240
-	10	DGGTTFEHLWSLSEPDSTYF	29
-	241	TCGACCTCCCCAGTCAGCCGGGAAATAATGAGGTGGTGGCACGGATTCCAGCA	300
-	30	DLPQSRSRNNEVVGTDSSM	49
-	301	TGGACGTCTCCACCTAGAGGGCATGACCACATCTGTCATGGCCAGTTCAATTGCTGA	360
-	50	DVFHLEGMTTSVMAQFNLLS	69
-	361	GCAGCACCATGGACCAAGATGAGCACGCCGCTGCCCTGGCCAGCCGTACACCCGGAGC	420
-	70	STMMDQMSSRAASASPYTPPEH	89
-	421	ACGCCGCAGCGTCCCCACCCATTACGCCAACAGCCAGCTCCACCTTCGACACCA	480
-	90	AASVPTHSPYAQOPSSTFDTM	109
-	481	TGTCGCCGCGCCTGTTCATCCCCCTCAACACCAGACTATCCGGACCCCACTTCGAGG	540
-	110	SPAPVIPSNTDYPGPHHFEV	129
-	541	TCACCTTACCACTCAGCACGCCAACGTCAGCCACCTGGACGTACTCCCACACTCTGAA	600
-	130	TFQQSSATAKSATWTSPLLLK	149
-	601	AGAAAATCTACTGCCAGATGCCAACAGCATGCCCATCCAGATCAAGGTGTCGCCAAC	660
-	150	KLYCQOIAKTCPIQIKVSAAPP	169
-	661	CGCCCCGGCACCGCATCCGGGATGCCGTACAAGAACGGGAGCACGTGACCG	720
-	170	PPGTAIRAMPVYKKAEHVTTD	189
-	721	ACATCGTAAGCGCTGCCAACACAGAGCTGGAGGGACTAACGAAGGACAGTCTG	780
-	190	IVKRCPNHELGDRDFNEGQSA	209
-	781	CCCCAGCCAGCCACCTCATCCGTGTTGAAGGCAATAATCTCTCCAGTATGTGGACGACC	840
-	210	PASHLIRVEGNLSQLQVDDP	229
-	841	CTGTACCCGGCAGGCAGGGCTCGTGGTGCCTATGAGCCACAGTGGGGACAGAAT	900
-	230	VTGRQSUVVPYEPPQVGTEF	249
-	901	TCACCACTCTGTACAACCTCATGTTAACAGCAGCTGTGTGGGGCATGAACCGAC	960
-	250	TTILYNFMCMNSCSVCGGMNRR	269
-	961	GGCCCATCTCATCATCACCCCTGGAGACGCCGGATGGCAGGTGCTGGCCGGCGGT	1020
-	270	PILIITLETRDGQVLGRRS	289
-	1021	CCTTCGAGGGCGCATCGCCCTGTCTGGCCGGACCGAAAAGCCGATGAGGACACT	1080
-	290	FEGRICACPCPRDRKADEDHY	309
-	1081	ACCGGGAGCAGCAGGCCCTGAATGAGAGCTCCGCAAGAACGGGCTGCCAGCAAGCGCG	1140
-	310	REQQALNESSSAKNGAASKRA	329
-	1141	CCTTCAGAGAGTCCCCCTGCCCTGGCCCTGGCCGGTGTGAAGAACGGCGGGC	1200
-	330	FKQSPPAVPALGPGVKKRRH	349
-	1201	ACGGAGACGAGGACACGTTACCTCGAGGTGGAGGCCGAGAACCTCGAGATCTGAA	1260
-	350	GDEDTYYLQVRGRENFEILM	369
-	1261	TGAAGCTGAAGGAGAGCCTGGAGCTGGAGTTGGTCCCGAGCCGCTGGTAGACTCCT	1320
-	370	KLKESLELHELVLVPQLVDSY	389
-	1321	ATCCGCAGCAGCACGCCAGCTCTACAGGCCAGCTACAGCCCCCATCTACGGGC	1380
-	390	RQQQQQLLQRPSYGP	409
-	1381	CGGTCTCTCGCCCATGAAACAGGTCACGGGGGGTGAACAAGCTGCCCTCCGTCAACC	1440
-	410	VLSPMNKVHGGVNKLPSSVNQ	429
-	1441	AGCTGGTGGGCCAGGCCCTCCCCCGCACGCTCGGCAGCTACACCCAACCTGGACCTGTGG	1500
-	430	LVGQPPPHSSATTPNLGPVVG	449
-	1501	GCTCTGGGATGCTCAACACCAGGCCACGCCAGTCAGCCACAGGAGATGACCGACA	1560
-	450	SGMLNNHGHAVPANSSEMSTS	469
-	1561	GCCACGGCACCCAGTCCATGGTCTGGGGTCCCACTGCACTCCGCAACCCCTACCCACG	1620
-	470	HGTQSMSGSHTCPPPPYPHA	489
-	1621	CCGACCCAGCCCTCGTCACTTTAACAGGATGGGTGTCAAACGTGCACTCGACTATT	1680
-	490	DPSLVSFLTGLGCPNCIEYF	509

FIG.4

8 / 36

-	1681	TCACGTCCCAGGGTTACAGAGCATTACCACTGAGAACCTGACCATCGAGGACCTGG	1740
-	510	T S Q G L Q S I Y H L Q N L T I E D L G	529
-	1741	GGCCCTGAAGATCCCCGAGCAGTATCGCATGACCATCTGGGGGGCTGCAGGACCTGA	1800
-	530	A L K I P E Q Y R M T I W R G L Q D L K	549
-	1801	AGCAGGGCCACGACTACGGGCCGCCGCCAGCAGCTGCTCCGCTCCAGAACGCCCG	1860
-	550	Q G H D Y G A A A Q Q L L R S S N A A A	569
-	1861	CCATTCCATGGGGCTCCGGGAGCTGCAGGCCAGCGGGCATGGAGGCCGTCACT	1920
-	570	I S I G G S G E L Q R V M E A V H F	589
-	1921	TCCCGTGCACACCATCACCATCCCCAACCGGGGCCGGGGGGGGGGGGGGGGGACG	1980
-	590	R V R H T I T I P N R G G P G A G P D E	609
-	1981	AGTGGGCGGACTTCGGCTTGCACCTGCCGACTGCAAGGCCGCAAGCAGCCCATCAAGG	2040
-	610	W A D F G F D L P D C K A R K Q P I K E	629
-	2041	AGGAGTTCACTGGAGGCCAGATCCACTGAGGGCCGGGGCCAGCAGGCCATGTGCCACC	2100
-	630	E F T B A E I H *	649
-	2101	GCCAGAGACCCAGGCCCTCGCTCTCCCTGTGTCCAAAATGCCCTCCGGAGCCAG	2160
-	2161	GGCTCCAGGCTGTGGGGAAAGGCAGGTCCGGCCATGCCCGCACCTCACCGG	2220
-	2221	CCCCAGGAGAGGCCAGGCCAACAGCCGGCTGGACAGCCTGAGTCACCTTGAGAACCC	2280
-	2281	TTCGGAGCTGCCCTAATGCTGGCTTGGGGCAGGGCCGGCCACTCTCAGCCCTGC	2340
-	2341	CACTGCCGGCGTGTCCATGGCAGGCCTGGGGACCGCAGTGTCAAGCTCCGACCTC	2400
-	2401	CAGGCCATCCTAGAGACTCTGTCAATCTGCCATCAAGCAAGTCCTTCCAGAGGAAG	2460
-	2461	AATCCCTTCGCTGGACTGCCAAAAAGTATTTGGACATCTTTGGTTCTGGAGAG	2520
-	2521	TGGTGGAGCAGCCAAAGCAGCTGTCTGAAACACCGTGCATTTCAAGGAATGCTCTAAC	2580
-	2581	GGGCTGGGACTCTCTGCTGGACTTGGAGTGGCTTGGCCCCAGCACACTGTATT	2640
-	2641	TGGGGGACCGCCCTCCCTGCCCTAACAAACCCAAAGTGTGCTGAAATTGGAGAAA	2700
-	2701	ACTGGGGAAAGGCCAACCCCTCCAGGTGGGGAGCATCTGGTACCCCTCGCCAGTG	2760
-	2761	CCCCTAGCCTGGCCACAGTCACCTCTCCCTGGGAACCTCTGGCAGAAAGGACAGCCT	2820
-	2821	GTCTTAGAGGACCGGAAATTGTCATATTTGATAAAATGATACCCCTTCTAC	2874

FIG.4 cont.

9 / 36

1	TGCCTCCCCGGCCGCA	CGACCCGGCCGAGGCC	TGTGCTCTGCGA	AGGGACGCAGCGAA	60
61	GCCGGGGCCCGCGC	CAGGCCGGCCGGACGG	ACGGCGATGCCG	AGCTGCGACGGCTGC	120
121	AGAGCGAGCTGCC	TCGGAGGCCGTGAGGA	AGATGGCCCAGTCC	ACCCACCTCCCC	180
-10			MAQ	STTTSP	9
181	CCGATGGGGCACCG	CGTMTGAGCACCT	CTGGAGCTCTCTG	GAACCAGACAGCACCTACT	240
10	DGGTTTFEHLWSS	LEPDSTYF			29
241	TCGACCTTCCCAGT	CAAGCCGGGAAATA	ATGAGGTGGTGGG	TGGCACGGATTCCAGCA	300
30	DLPQSRRGN	NEVVG	GTDSM		49
301	TGGACGTCTTCCAC	CTAGAGGGCATGACCACAT	CTGTCATGGCCCAGT	TCAATTGCTGA	360
50	DVFHLEGMTTSV	HMAQFNL	LSS		69
361	GCAGCACCATGGACC	AGATGAGCAGGCC	CGCTCGGCC	AGGCCGCTACACCCCGAGC	420
70	STMDDQMSSRA	ASASPY	TPEH		89
421	ACGCCGCCAGCGT	GCCCCACCCATT	ACCCCTACCGCACAG	CCGCTCCACCTTCGACACCA	480
90	AASVPTHSPYA	QPPS	STFD	DTM	109
481	TCTGCCCGCCGCTG	TATCCCCTCAACACCG	ACTATCCC	GGACCCACCACTTGAGG	540
110	SPA	PVIPSNTD	YPGPHF	EVE	129
541	TCAC	TTCCAGCAGTC	CCAGCACGGCCA	AGTCAGCCACCTGGAC	600
130	T	FQQS	STA	WTY	SPPLK
601	AGAA	ACTCTACTGCCAG	ATGCCAAGACATGCC	CCATCCAGATCAAGGTGT	660
150	KLYCQIAKT	TCPI	QIKVS	APP	169
661	CGCCCCCGGG	ACCGCCATCGGCC	ATCGCTGCTACA	AGAAGCCGAGCACGTGACCG	720
170	PPGT	TAIRAM	PVYK	RAEHVT	189
721	ACATCGTGAAGCG	CTGCCAAC	ACAGAGCTGGGAGGG	ACTTCAACGAAGGACAGTCG	780
190	I	VKRCPNHE	LGRDF	NEGQSA	209
781	CCCCAGCCAGC	ACCCATCGTGTG	GAAGCAATA	ATCTCGCAGTATGTGGACGACC	840
210	PASHLIR	VEG	GNNLSQY	VVD	229
841	CTGTCACCGG	CAGGCAGAGCG	CTGTGGTGCC	CTATGAGCCAC	900
230	V	TGROS	VVP	YEP	249
901	TCAC	ACCCATCTGTA	CAACTCATG	GTCAGCAGCTGTGTGGGG	960
250	TTI	LYNFMC	NSSCV	GGMNR	269
961	GGCCC	ATCCTCATCAT	CACCC	TGGAGACGCCGGATGGCAGGTG	1020
270	PILLI	IIITL	ETR	RDGQVLGR	289
1021	CCTTC	GAGGGCCG	CATCTGCGC	CTGTC	1080
290	FEGRI	ACPGR	DRDKA	DEDEDHY	309
1081	ACCGGGAGCAGC	AGCCCTTG	AAATGAGAC	GTCCGCCAACGACGGG	1140
310	REEQ	QALNES	SAKNGA	ASKRA	329
1141	CCTTC	AAAGCAGT	CCCCCTGCG	TCCCCCGCTGGGCGAGTGTG	1200
330	F	KQSP	VAP	PLGPVVKR	349
1201	ACGGAGACGAGG	ACCGTACTAC	CTGAGGTGCG	GAGGCCGAGAAC	1260
350	G	DED	TYLQVR	GRRENFEIL	369
1261	TGAAGCT	GAAGGAGAC	CTGGAGCTGAT	GGAGTTGGTGCCG	1320
370	K	LKE	S	ELMELV	389
1321	ATCGGCAGCAGC	AGCAGCTCT	ACAGAGGCC	GAGTCACCTACAG	1380
390	R	QQQQL	L	QRP	409
1381	CGGT	CTCTGCC	CATGAAC	AGGTGCACGGGGCG	1440
410	V	LSPMN	KV	HGGV	429
1441	AGCTGGTGGCC	CAGCC	CTCCCCG	ACAGCTGGCAGCTAC	1500
430	L	VGQ	PPP	PHS	449
1501	GCTCTGGG	ATGCTCA	ACACACC	GGCCACGGCAG	1560
450	S	GM	LNN	HGA	469
1561	GCCACGG	CAACCC	AGCTCATG	GTCTCGCC	1620
470	H	GTQ	SMV	SGSH	489
1621	CCGAC	CCCCAGC	CTCGTC	AGGACCTGGGG	1680
490	D	PSL	VRT	WGP	509
1681	CATCTGG	GGGGCTCG	AGGACCTG	GGACGACTACGG	1740
1741	GCTG	CTCCGCT	CAGCAAC	GGCCGCGC	1800
1801	CCAGCGGG	TCATGGAG	CCGTG	ACTTC	1860
1861	CGGCGG	CCCCGG	GGCCGG	GGAGACTCGG	1920
1921	CAAGG	CCCCGCA	AGCAG	CCATCAAGGAGG	1980
1981	CGGG	CCCCAGCC	AGAGC	CTGTGCC	2034

FIG.5

10/36

1	GCGAGCTGCCCTCGGAGGCCGGCGTGGGAAGATGGCCCAGTCCACCGCCACCTCCCTG	60
-9	M A Q S T A T S P D	10
61	ATGGGGGCACCACGTTGAGCACCTCTGGAGCTCTGGAACCCAGACAGCACCTACTTCG	120
11	G G T T F E H L W S S L E P D S T Y F D	30
121	ACCTCCCCAGTCAAGCCGGGAAATAATGAGGTGGTGGCGGAACGGATTCCAGCATGG	180
31	L P Q S S R G N N E V V G G T D S S M D	50
181	ACGTCTTCCACCTGGAGGGCATGACTACATCTGTCAATGGCCCAGTCAATCTGCTGAGCA	240
51	V F H L E G M T T S V M A Q F N L L S S	70
241	GCACCATGGACCAGATGAGCAGCCGCGGCCCTGGCCAGCCCCAACCCCCAGAGCACG	300
71	T M D Q M S S R A A S A S P Y T P E H A	90
301	CCGCCAGCGTGCCAACCCACTGCCCTACGCACAACCCAGCTCCACCTTCGACACCATGT	360
91	A S V P T H S P Y A Q P S S T F D T M S	110
361	CGCCGGCGCCTGTCATCCCCCTCCAACACCGACTACCCGGACCCACCACCTTGAGGTCA	420
111	P A P V I P S N T D Y P G P H H F E V T	130
421	CTTTCAGCAGTCCAGCACGGCCAAGTCAGCCACCTGGACGTACTCCCCGCTCTGAAGA	480
131	F Q Q S S T A K S A T W T Y S P L L K K	150
481	AACTCTACTGCCAGATGCCAACAGACATGCCCATCCAGATCAAGGTGTCCACCCGCCAC	540
151	L Y C Q I A K T C P I Q I K V S T P P P	170
541	CCCCAGGCACTGCCATCCGGGCCATGCTGTAAAGAAAAGCGGAGCACGTGACCGACG	600
171	P G T A I R A M P V Y K K A E H V T D V	190
601	TCGTAAACGCTGCCAACCACGAGCTCGGGAGGACTCAACGAAGGACAGTCTGCTC	660
191	V K R C P N H E L G R D F N E G Q S A P	210
661	CAGCCAGCCACCTCATCCGCGTGGAAAGCAATAATCTCGCAGTATGTGGATGACCCCTG	720
211	A S H L I R V E G N N L S Q Y V D D P V	230
721	TCACCGCAGGCAGAGCGTCGTGGTGCCTATGAGCCACCACAGGTGGGACGGAATTCA	780
231	T G R Q S V V V P Y E P P Q V G T E F T	250
781	CCACCATCCTGTACAACCTCATGTGTAAACAGCAGCTGTAGGGGCATGAACCGGCGGC	840
251	T I L Y N F M C N S S C V G G M N R R P	270
841	CCATCCTCATCATCACCTGGAGATGGGGATGGCAGGTGCTGGGCCGGTCCT	900
271	I L I I I T L E M R D G Q V L G R R S F	290
901	TTGAGGGCCGCATCTGCGCCTGTCCCTGGCGCGACCCAAAAGCTGATGAGGACCACTACC	960
291	E G R I C A C P G R D R K A D E D H Y R	310
961	GGGAGCAGCAGGCCCTGAACGAGAGCTCCGCCAAGAACGGGCCAGCAAGCGTGCCT	1020
311	E Q Q A L N E S S A K N G A A S K R A F	330
1021	TCAAGCAGAGCCCCCTGCCGTCCCCGCCCTGGTGCCTGGTGTGAAGAACGGCGGCATG	1080
331	K Q S P P A V P A L G A G V K K R R H G	350
1081	GAGACGAGGACACGTACTACCTTCAGGTGCGAGGCCGGAGAACCTTGAGATCCTGATGA	1140
351	D E D T Y Y L Q V R G R E N F E I L M K	370
1141	AGCTGAAAGAGAGCTGGAGCTGATGGAGTTGGTGCCTGGCAGCCACTGGTGGACTCTATC	1200
371	L K E S L E L M E L V P Q P L V D S Y R	390
1201	GGCAGCAGCAGCAGCTCCTACAGAGGCCAGTCACCTACAGCCCCCTACGGGCCGG	1260
391	Q Q Q Q L L Q R P S H L Q P P S Y G P V	410
1261	TCCTCTGCCCATGAACAAGGTGCACGGGGCATGAACAAGCTGCCCTCCGTCAACCGC	1320
411	L S P M N K V H G G M N K L P S V N Q L	430
1321	TGGTGGGCCAGCCTCCCCCGCACAGTCGGCAGCTACACCCAACTGGGCCGTGGCC	1380
431	V G Q P P P H S S A A T P N L G P V G P	450
1381	CCGGGATGCTCAACAACCATGGCCACGCCAGCAACGGCAGATGAGCAGCAGCC	1440
451	G M L N N H G H A V P A N G E M S S S H	470

FIG.6

11 / 36

1441	ACAGCGCCCAGTCCATGGTCTCGGGGTCCCCTGCACACTCCGCCACCCCCCTACCACGCCG	1500
471	S A Q S M V S G S H C T P P P P Y H A D	490
1501	ACCCCAGCCTCGTCAGTTTTAACAGGATTGGGGTGTCAAACGTGCATCGAGTATTCA	1560
491	P S L V S F L T G L G C P N C I E Y F T	510
1561	CCTCCCAAGGGTTACAGAGCATTTACCCACCTGCAGAACCTGACCATTGAGGACCTGGGG	1620
511	S Q G L Q S I Y H L Q N L T I E D L G A	530
1621	CCCTGAAGATCCCCGAGCAGTACCGCATGACCATCTGGGGGCTGCAGGACCTGAAGC	1680
531	L K I P E Q Y R M T I W R G L Q D L K Q	550
1681	AGGGCCACGACTACAGCACCGCGCAGCAGCTGCTCCGCTTAGCAACCGGGCCACCATCT	1740
551	G H D Y S T A Q Q L L R S S N A A T I S	570
1741	CCATGGCGGCTCAGGGAACTGCAGGCCAGGGTCATGGAGGCGTGACTTCCGCG	1800
571	I G G S G E L Q R Q R V M E A V H F R V	590
1801	TGCGCCACACCATCACCATCCCCAACCGCGGCGGCCAGGGCGGCCCTGACGAGTGGG	1860
591	R H T I T I P N R G G P G G G P D E W A	610
1861	CGGACTTCGGCTTCGACCTGCCGACTGCAAGGCCGCAAGCAGCCATCAAGGAGGAGT	1920
611	D F G F D L P D C K A R K Q P I K E E F	630
1921	TCACGGAGGCCGAGATCCACTGAGGGCCTCGCCTGGCTGCAGCCTGCCACCGCCAGA	1980
631	T E A E I H *	650
1981	GACCCAAGCTGCCCTCCCTCTCCTCTGTGTGTCCAAAACGCCCTCAGGAGGCAGGACC	2040
2041	TTCGGGCTGTGCCGGGAAAGGCAAGGTCCGCCATCCCCAGGCACCTCACAGGCC	2100
2101	AGGAAAGGCCAGCCACCGAAGCCCTGTGGACAGCCTGAGTCACCTGCAGAAC	2156

FIG. 6 cont.

12 / 36

1 TGATCTCCCTGTGGCTGCAAGGGACTGACCCAGGGAGTAGATGCCCTGAGACCCAAAGG
 61 GACACCAAGGAACCTTGTGGCTTGAAAGGGATCGTCCTCTCTGCCAAGAGA 60
 121 AGCATGTGTATGGGCCCTGTATGAATCCTGGGGCAGGCCAGTCATTGTCAGC 120
 0 M C M G P V Y E S L G Q A Q P N L L S 180
 181 AGTGCATGGACCAAGATGGCAGCCGTCCCCCCCCGGAGCCCTACACCCGGAC 19
 20 S A M D Q M G S R A A P A S P Y T P E H 240
 241 GCGCCAGCGCGCCACCCACTCGCCATCGCAGGCCAGTCACCTCGACACC 39
 40 A A S P T H S P Y A Q P S S T F D T M 300
 301 TCTCCGGCGCTGTCACTCCCTCCAATACCGACTACCCGGCCACCTTCGAGG 59
 60 S P A P V I P S N T D Y P G P H F E V 360
 361 ACCTTCCAGCAGTCGAGCACTGCCAAGTCGGCCACCTGGACATACTCCC 79
 80 T F Q Q S S T A K S A T W T Y S P L L K 420
 421 AAGTTGTACTGTCAAGACATGCCCATCCAGATCAAAGTGTCCACACCACCA 99
 100 K L Y C Q I A K T C P I Q I K V S T P P 480
 481 CCCCCGGGACGGCCATCCGGGCGATGCCGTGTCAGAAGAGGCAAGACATGTGACCGAC 119
 120 P P G T A I R A M P V Y K K A E H V T D 540
 541 ATTGTAAAGCGCTGCCCAACCCAGAGCTGGAGGGACTTCATGAAGGACAGTGTGCC 139
 140 I V K R C P N H E L G R D F N E G Q S A 600
 601 CCGGCTAGCCACCTCATCCGTGAGAAGGCAACAACCTGCCAGTACGTGGATGACCC 159
 160 P A S H L I R V E G N N L A Q Y V D D P 660
 661 GTCAACCGAAGGCAGAGTGTGTTGTGCCATGAACCCCCACAGTGGAACAGAA 179
 180 V T G R Q S V V V P Y E P P Q V G T E F 720
 721 ACCACCATCTGTACAACCTCATGTGTAACAGCAGCTGTGAGGGGCAATGAATCGGAGG 199
 200 T T I L Y N F M C N S C V G G M N R R 780
 781 CCCATCTGTCACTCATCACCTGGAGACCCGGATGGACAGCTCTGGGCCCGCT 219
 220 P I L V I T L E T R D G Q V L G R R S 840
 841 TTGAGGGCTCGCATCTGTGCCCTGCTGGCGAGCTGACAGCTGATGAAGACCA 239
 240 F E G R I C A C P G R D R K A D E D H Y 900
 901 CGGGAGCAACAGGCTCTGAATGAAAGTACCAACAAAATGGAGCTGCCAGCAACGTGCA 259
 260 R E Q O A L N E S T T K N G A A S K R A 960
 961 TTCAAGCAGACCCCCCTGCCATCCCTGGGTACCAACAGTGAAGAAGAGACCC 279
 280 F K Q S P P A I P A L G T N V K K R R H 1020
 1021 GGGGACGAGGACATGTTCTACATGCACCTGCGAGGGGGAGAACCTTGAGATCTTGATG 299
 300 G D E D M F Y M H V R G R E N F E I L M 1080
 1081 AAAGTCAGGGAGGAGCTAGAACCTGAGCTGGAGCTGTGCCACCCCTGGTGA 319
 320 K V K E S L E L M V P Q P L V D S Y 1140
 1141 CGACAGCAGCAGCAGCAGCAGCTCTACAGAGGCCAGTCACCTGCAGCCCTCATCT 339
 340 R Q Q Q Q Q L L Q R P S H L Q P P S Y 1200
 1201 GGGCCCGTGTCTCCCAATGAACAGGTACACGGTGGTGTCAACAACTGCCCTCCGT 359
 360 G P V L S P M N K V H G G V N K L P S V 1260
 1261 AACCACTGGTGGGCCAGCCCTCCCCACAGCTCAGCAGCTGGGCAACCTGGGCC 379
 380 N Q L V G Q P P P H S S A A G P N L G P 1320
 1321 ATGGGCTCCGGATGTCACAGCCACGGCCACAGCATGCCGCCATGGTGA 399
 400 M G S G M L N S H G H S M P A N G E M N 1380
 1381 GGAGGCCACAGCTCCAGACCATGGTTGGGACTGCCACTGCAACCCGCCACCCCT 419
 420 G G H S S Q T M V S G S H C T P P P Y 1440
 1441 CATGAGACCCCCAGCTCGTCAGTTTGACAGGGTTGGGTGTCACACTGCA 439
 440 H A D P S L V S F T L G L G C P N C I E 1500
 1501 TGCTTCACCTCCCAAGGGTTGAGAGCATCTACCCACCTGCCAGAACCTTAC 459
 460 C F T S Q G L Q S I Y H L Q N L T I E D 1560
 1561 CTGGGGCTCTGAAGGTCTGACCACTGACCTGACGGTATGACCACTGGAGGGCT 479
 480 L G A L K V P D Q Y R M T I W R G L Q D 1620
 1621 CTGAAGCAGAGCCATGACTGCCAGCACTGCTACGCTCCAGCAGCAACGCC 499
 500 L K Q S H D C G Q O L R S S S N A A T 1680
 1681 ATCTCCATCGGGGGCTCTGGCGAGCTGCGAGCCGAGCGGGTCATGGAAGCC 519
 520 I S I G G S G E L Q R Q R V M E A V H F 1740
 1741 COTGTGGCCACACCATCACAAATCCCCAACCGTGGAGGCCAGGTGGCGGT 539
 540 R V R H T I T I P N R G G A G A V T G P 1800
 1801 GACGAGTGGGGACTTTGGCTTGACCTGCCCTGACTGCAAGTCCCCTAAC 559
 560 D E W A D F G F D L P D C K S R K Q P I 1860
 1861 AAAGAGGAGGTACAGAGACAGAGGCCACTGAGGAACGTACCTCTCTGTCTTC 579
 580 K E E F T E T E S H * 1920
 1921 CTCTGTGAGAAACTGCTCTTGGAGGGACTGTTGGCTGTGCCCCACAGAAAC 599
 1981 GGACCTCTGCCGGATGCCATTCTGAAGGGAGTCGCTCATGAACACTCCC 1980
 1981 TTGG 2040

FIG.7

13 / 36

-	-	1	TGGTCCCCTCGACCAAGACTCCGGTACCAAGCTTGCGGCCCCCGCGGAGGGAGGAGACC	60
-	-	61	CCGCTGGGCTAGCTGGCGACGCGGCCAAGCGGCCGGGAAGGAGCCGGAGGAGCG	120
-	-	121	GGGCCCAGACCCCCACTCGGGCAGAGCCAGCTGGGAGGCAGGGCGCCGTGGAGCCA	180
-	-	181	GGGGCCCGGGTGGCCGCCCTCCCTCCGCCACGGCTGAGTGCGCTGCCCTCCGCCG	240
-	-	241	GTCGGCAAGAACGGCTAACGCTCGGCAGTCCCCTCGCCGCCCTCCGTCTCCGC	300
-	-	301	ACCCATTATAACCCGGCTCCCGCATCCAGGCAGGAGCAACGCTGCAAGCCAGCCCTCG	360
-	-	361	CCGACGCCGACGCCGGCCGGAGCAGAATGAGCGGCAGCGTTGGGAGATGGCCAGAC	420
-	-	-8	M S G V G E M A Q T	11
-	-	421	CTCTTCTCTCTCCACCTCGAGCACCTGTGGACTCTCTAGAGCCAGACAGCAC	480
-	-	12	S S S S S T F E H L W S S L E P D S T	31
-	-	481	CTACTTTGACCTCCCCAGCCAGCCAAGGGACTAGCGAGGCATCAGGCAGCGAGGAGTC	540
-	-	32	Y F D L P Q P S Q G T S E A S G S E E S	51
-	-	541	CAACATGGATGTCTTCCACCTGCAAGGCATGGCCAGTTCAATTGCTCAGCAGTGCAT	600
-	-	52	N M D V F H L Q G M A Q F N L L S S A M	71
-	-	601	GGACCAAGATGGCACCCGTGCGGGCCCGCGAGCCCTACACCCGGACCGCCAGCAG	660
-	-	72	D Q M G S R A A P A S P Y T P E H A A S	91
-	-	661	CGCGCCACCCACTCGCCCTACGCCAGCCAGCTCCACCTCGACACCATGTCTCCGGC	720
-	-	92	A P T H S P Y A Q P S T F D T M S P A	111
-	-	721	GCCTGTCATCCCTCCAATACCGACTACCCGGCCCCC	758
-	-	112	P V I P S N T D Y P G P	123

FIG. 8

```

- Name: sr-p70a-cos3 Len: 650 Check: 9661 Weight: 1.00
- Name: sr-p70b-cos3 Len: 650 Check: 3605 Weight: 1.00
- Name: sr-p70-ht29 Len: 650 Check: 85 Weight: 1.00
- Name: sr-p70c-att20 Len: 650 Check: 4072 Weight: 1.00
- Name: sr-p70a-att20 Len: 650 Check: 4204 Weight: 1.00

- //
-
- 1
- sr-p70a-cos3 ..... MAQ STTTSPDGGT TFEHLWSSLE PDSTYFDLPQ SSRGNNEVVG 50
- sr-p70b-cos3 ..... MAQ STTTSPDGGT TFEHLWSSLE PDSTYFDLPQ SSRGNNEVVG
- sr-p70-ht29 ..... MAQ STATSPDGGT TFEHLWSSLE PDSTYFDLPQ SSRGNNEVVG
- sr-p70c-att20 ..... MSGSVGEMAQ ... TSSSSSS TFEHLWSSLE PDSTYFDLPQ PSQGTSEASG
- sr-p70a-att20

- 51
- sr-p70a-cos3 GTDSSMD.VF HLEGMTTSVM AQFNLLSSTM DQMSSRAASA SPYTPHEAAS 100
- sr-p70b-cos3 GTDSSMD.VF HLEGMTTSVM AQFNLLSSTM DQMSSRAASA SPYTPHEAAS
- sr-p70-ht29 GTDSSMD.VF HLEGMTTSVM AQFNLLSSTM DQMSSRAASA SPYTPHEAAS
- sr-p70c-att20 ... MCMGPVY .. ESLG...Q AQFNLLSSAM DQMGSRAAPA SPYTPHEAAS
- sr-p70a-att20 SEESNMD.VF HLQGM.... AQFNLLSSAM DQMGSRAAPA SPYTPHEAAS

- 101
- sr-p70a-cos3 VPTHSPYAQP SSTFDTMSPA PVIPSNTDYP GPHHFEVTFQ QSSTAKSATW
- sr-p70b-cos3 VPTHSPYAQP SSTFDTMSPA PVIPSNTDYP GPHHFEVTFQ QSSTAKSATW
- sr-p70-ht29 VPTHSPYAQP SSTFDTMSPA PVIPSNTDYP GPHHFEVTFQ QSSTAKSATW
- sr-p70c-att20 APTHSPYAQP SSTFDTMSPA PVIPSNTDYP GPHHFEVTFQ QSSTAKSATW
- sr-p70a-att20 APTHSPYAQP SSTFDTMSPA PVIPSNTDYP GP..... 150

- 151
- sr-p70a-cos3 TYSPLLKKLY CQIAKTCPIQ IKVSAPPPP TAIRAMPVYK KAEHVTDIVK 200
- sr-p70b-cos3 TYSPLLKKLY CQIAKTCPIQ IKVSAPPPP TAIRAMPVYK KAEHVTDIVK
- sr-p70-ht29 TYSPLLKKLY CQIAKTCPIQ IKVSTPPPP TAIRAMPVYK KAEHVTDIVK
- sr-p70c-att20 TYSPLLKKLY CQIAKTCPIQ IKVSTPPPP TAIRAMPVYK KAEHVTDIVK
- sr-p70a-att20

- 201
- sr-p70a-cos3 RCPNHELGRD FNEGQSAPAS HLIRVEGNL SQYVDDPVTG RQSVVVPYEP 250
- sr-p70b-cos3 RCPNHELGRD FNEGQSAPAS HLIRVEGNL SQYVDDPVTG RQSVVVPYEP
- sr-p70-ht29 RCPNHELGRD FNEGQSAPAS HLIRVEGNL SQYVDDPVTG RQSVVVPYEP
- sr-p70c-att20 RCPNHELGRD FNEGQSAPAS HLIRVEGNL AQYVDDPVTG RQSVVVPYEP
- sr-p70a-att20

- 251
- sr-p70a-cos3 PQVGTEFTTI LYNFMCNSSC VGGMNRRL PIL IIITLETRDG QVLGRRSFEG 300
- sr-p70b-cos3 PQVGTEFTTI LYNFMCNSSC VGGMNRRL PIL IIITLETRDG QVLGRRSFEG
- sr-p70-ht29 PQVGTEFTTI LYNFMCNSSC VGGMNRRL PIL IIITLEMRDG QVLGRRSFEG
- sr-p70c-att20 PQVGTEFTTI LYNFMCNSSC VGGMNRRL PIL VIITLETRDG QVLGRRSFEG
- sr-p70a-att20

- 301
- sr-p70a-cos3 RICACPGRDR KADEDHYREQ QALNESSAKN GAASKRAFKQ SPPAVPALGP 350
- sr-p70b-cos3 RICACPGRDR KADEDHYREQ QALNESSAKN GAASKRAFKQ SPPAVPALGP
- sr-p70-ht29 RICACPGRDR KADEDHYREQ QALNESSAKN GAASKRAFKQ SPPAVPALGA
- sr-p70c-att20 RICACPGRDR KADEDHYREQ QALNESTTKN GAASKRAFKQ SPPAIPALGT
- sr-p70a-att20
-
```

...

FIG.9

- sr-p70a-cos3 351 400
 - sr-p70b-cos3 GVKKRRHGDE DTYYLQVRGR ENFEILMKLK ESLELMELVP QPLVDSYR..
 - sr-p70-ht29 GVKKRRHGDE DTYYLQVRGR ENFEILMKLK ESLELMELVP QPLVDSYR..
 - sr-p70c-att20 NVKKRRHGDE DMFYMHVRGR ENFEILMKVK ESLELMELVP QPLVDSYRQQ
 - sr-p70a-att20

 - sr-p70a-cos3 401 450
 - sr-p70b-cos3 QQQQLLQRPS HLQPPSYGPV LSPMNKVHGG VNKLPSVNQL VGQPPPHSSA
 - sr-p70-ht29 QQQQLLQRPS HLQPPSYGPV LSPMNKVHGG VNKLPSVNQL VGQPPPHSSA
 - sr-p70c-att20 QQQQLLQRPS HLQPPSYGPV LSPMNKVHGG MNKLPSVNQL VGQPPPHSSA
 - sr-p70a-att20

 - sr-p70a-cos3 451 500
 - sr-p70b-cos3 ATPNLGPVGS GMLNNGHAV PANSEMTSSH GTQSMVSGSH CTPPPPYHAD
 - sr-p70-ht29 ATPNLGPVGS GMLNNGHAV PANSEMTSSH GTQSMVSGSH CTPPPPYHAD
 - sr-p70c-att20 ATPNLGPVGP GMLNNGHAV PANGEMSSH SAQSMVSGSH CTPPPPYHAD
 - sr-p70a-att20 AGPNLGPMS GMLNSHGHSN PANGEMNGGH SSQTMVSGSH CTPPPPYHAD

 - sr-p70a-cos3 501 550
 - sr-p70b-cos3 PSLVSFLTGL GCPNCIEYFT SQGLQSIYHL QNLTIEDLGA LKIPEQYRMT
 - sr-p70-ht29 PSLVR.. T.W.G.P.....
 - sr-p70c-att20 PSLVSFLTGL GCPNCIEYFT SQGLQSIYHL QNLTIEDLGA LKIPEQYRMT
 - sr-p70a-att20 PSLVSFLTGL GCPNCIECFT SQGLQSIYHL QNLTIEDLGA LKVPDQYRMT

 - sr-p70a-cos3 551 600
 - sr-p70b-cos3 IWRGLQDLKQ GHDYGAAAQQ LLR.SSNAAA ISIGGSGELO RQRVMEAVHF
 - sr-p70-ht29 IWRGLQDLKQ GHDYS.TAQQ LLR.SSNAAT ISIGGSGELO RQRVMEAVHF
 - sr-p70c-att20 IWRGLQDLKQ SHDCG...QQ LLRSSSSNAAT ISIGGSGELO RQRVMEAVHF
 - sr-p70a-att20

 - sr-p70a-cos3 601 650
 - sr-p70b-cos3 RVRHTITIPN RGGPGA..GP DEWADFGFDL PDCKARKQPI KEEFTEAEIH
 - sr-p70-ht29 RVRHTITIPN RGGPGG..GP DEWADFGFDL PDCKARKQPI KEEFTEAEIH
 - sr-p70c-att20 RVRHTITIPN RGGAGAVTGP DEWADFGFDL PDCKSRKQPI KEEFTETESH
 - sr-p70a-att20

 -

 -

FIG.9 cont.

16/36

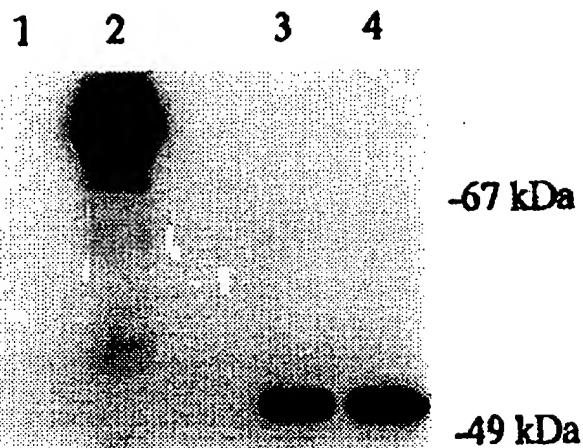


FIG.10a

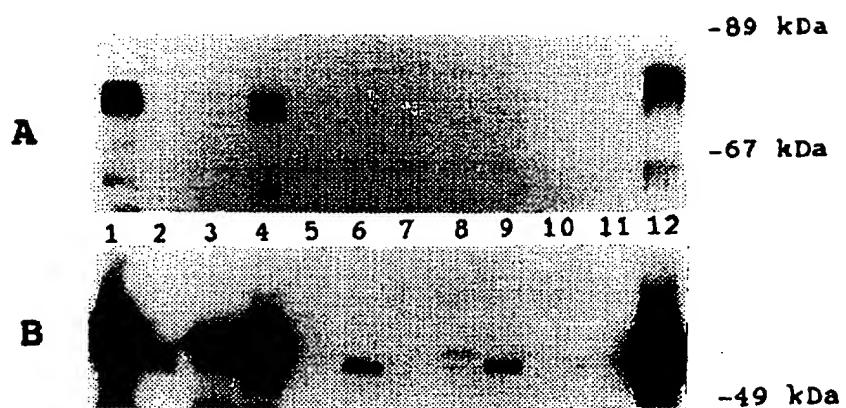


FIG.10 b

17/36



FIG.11

18 / 36

FIG.12

20/36

sr-p70d-imr32	CG ACCTTCCCCA GTCAAGCCGG GGGATAATG 32
sr-p70a-ht29	CG ACCTTCCCCA GTCAAGCCGG GGGATAATG 150
	AGGTGGTGGG CGGAACGGAT TCCAGCATGG ACGTCTTCCA CCTGGAGGGC 82
	AGGTGGTGGG CGGAACGGAT TCCAGCATGG ACGTCTTCCA CCTGGAGGGC 200
	ATGACTACAT CTGTCATGCA TCCTCGGCTC CTGCCTCACT AGCTGGGAG 132
	ATGACTACAT CTGTCAT..... 217
	CCTCTCCCAC TCGGTCCACG CTGCCGGCG GCCACGACCG TGACCCTTCC 182

	CCTCGGGCCG CCCAGATCCA TGCCTCGTCC CACGGGACAC CAGTTCCCTG 232

	GCGTGTGCAG ACCCCCCGGC GCCTACCATG CTGTACGTCG GTGACCCCGC 282

	ACGGCACCTC GCCACGGCCC AGTTCAATCT GCTGAGCAGC ACCATGGACC 332
 GGGCC AGTTCAATCT GCTGAGCAGC ACCATGGACC 252
	AGATGAGCAG CGCGCGGGCC TCGGCCAGCC CCTACACCCC AGAGCACGCC 382
	AGATGAGCAG CGCGCGGGCC TCGGCCAGCC CCTACACCCC AGAGCACGCC 302
	GCCAGCGTGC CCACCCACTC GCCCTACGCA CAACCCAGCT CCACCTTCGA 432
	GCCAGCGTGC CCACCCACTC GCCCTACGCA CAACCCAGCT CCACCTTCGA 352
	CACCATGTCG CGGGCGCCTG TCATCCCCCTC CAACACCGAC TACCCCGGAC 482
	CACCATGTCG CGGGCGCCTG TCATCCCCCTC CAACACCGAC TACCCCGGAC 402
	CCCACCACTT TGAGGTCACT TTCCAGCAGT CCAGCACGGC CAAGTCAGCC 532
	CCCACCACTT TGAGGTCACT TTCCAGCAGT CCAGCACGGC CAAGTCAGCC 452
	ACCTGGACGT ACTCCCCGCT CTTGAAG
	ACCTGGACGT ACTCCCCGCT CTTGAAG

FIG. 14

21/36

sr-p70a	T A A C G G C C C G G C C T A C T C C C G G C C T C C C C T C C C C C G G C C C A	50
sr-p70f	- - - - -	0
sr-p70d	- - - - -	0
sr-p70e	- - - - -	0
sr-p70b	- - - - -	0
sr-p70a	T A T A A C C C G G C C T A G G G G C C G C A G G C C T G C C C G C C C G G C A	100
sr-p70f	- - - - -	0
sr-p70d	- - - - -	0
sr-p70e	- - - - -	0
sr-p70b	- - - - -	0
sr-p70a	C C C G C C C G G A G G G G A C G C A G C G G A A C C G G G C	150
sr-p70f	- - - - -	0
sr-p70d	- - - - -	0
sr-p70e	- - - - -	0
sr-p70b	- - - - -	0
sr-p70a	C C C G C C C A G G C G G A C G G G C C G A	200
sr-p70f	T G C C C G G G G C T G C G A C G G C T	24
sr-p70d	T G C C C G G G G C T G C G A C G G C T	0
sr-p70e	- - - - -	0
sr-p70b	- - - - -	0
sr-p70a	G C A G A G C G A G C T G C C C G G C C G A G C G G G A A G T G G C C C A G T C C A	250
sr-p70f	G C A G - - - - -	0
sr-p70d	- - - - -	0
sr-p70e	- - - - -	0
sr-p70b	- - - - -	0

FIG. 15

22/36

sr-p70a	CCGCCACCTCCCCCTGATGGGGCACCAACCGTTGAGCACCCTCTGGAGCTCT	300
sr-p70f	-	24
sr-p70d	-	0
sr-p70e	-	0
sr-p70b	CCGCCACCTCCCCCTGATGGGGCACCAACCGTTGAGCACCCTCTGGAGCTCT	63
sr-p70a	CTGGAAACCAAGAACCTACTTCCGACCCATTCCGACCC	350
sr-p70f	-GGAAACCAAGAACCTACTTCCGACCCATTCCGACCC	72
sr-p70d	-	0
sr-p70e	-	0
sr-p70b	CTGGAAACCAAGAACCTACTTCCGACCCATTCCGACCC	113
sr-p70a	TAAATGAGGGTGGGTGGGAAACGGGATTCAGGATTCAGCTTCCACCTGG	400
sr-p70f	TAAATGAGGGTGGGTGGGAAACGGGATTCAGGATTCAGCTTCCACCTGG	122
sr-p70d	-ATGCCATGCTA[CCTGGTGAACCCGGGCAACCTTCCACCTGG	33
sr-p70e	-ATGCCATGCTA[CCTGGTGAACCCGGGCAACCTTCCACCTGG	33
sr-p70b	TAAATGAGGGTGGGTGGGAAACGGGATTCAGGATTCAGCTTCCACCTGG	163
sr-p70a	AGGGCATGACTACATCTGTCATGGCCCAGTTCAATCTGCTGAGCAGCAC	450
sr-p70f	AGGGCATGACTACATCTGTCATGGCCCAGTTCAATCTGCTGAGCAGCAC	172
sr-p70d	-GCCACATGCTA[CACGGGTGAACCCGGGCAACCTTCCACCTGG	66
sr-p70e	-GCCACATGCTA[CACGGGTGAACCCGGGCAACCTTCCACCTGG	66
sr-p70b	AGGGCATGACTACATCTGTCATGGCCCAGTTCAATCTGCTGAGCAGCAC	213
sr-p70a	ATGGACCAAGATGAGCAGCCAGCCATGGCTCCAGCTTACACCCCAAGA	500
sr-p70f	ATGGACCAAGATGAGCAGCCAGCCATGGCTCCAGCTTACACCCCAAGA	222
sr-p70d	ATGGACCAAGATGAGCAGCCAGCCATGGCTCCAGCTTACACCCCAAGA	116
sr-p70e	ATGGACCAAGATGAGCAGCCAGCCATGGCTCCAGCTTACACCCCAAGA	116
sr-p70b	ATGGACCAAGATGAGCAGCCAGCCATGGCTCCAGCTTACACCCCAAGA	263

FIG.15 cont.

23 / 36

sr-p70a	GCACCGCCAGCGTGGCCACCCACTCGCCCTACGCCACAA	550
sr-p70f	GCAAGCCAGCGTGGCCACCCACTCGCCCTACGCCACAA	272
sr-p70d	GCAAGCCAGCGTGGCCACCCACTCGCCCTACGCCACAA	166
sr-p70e	GCAAGCCAGCGTGGCCACCCACTCGCCCTACGCCACAA	166
sr-p70b	GCAAGCCAGCGTGGCCACCCACTCGCCCTACGCCACAA	313
sr-p70a	CCTTCGACACCATGTCGGCCGGCCTGTCAATCCCACACCGACTAC	600
sr-p70f	CCTTCGACACCATGTCGGCCGGCCTGTCAATCCCACACCGACTAC	322
sr-p70d	CCTTCGACACCATGTCGGCCGGCCTGTCAATCCCACACCGACTAC	216
sr-p70e	CCTTCGACACCATGTCGGCCGGCCTGTCAATCCCACACCGACTAC	216
sr-p70b	CCTTCGACACCATGTCGGCCGGCCTGTCAATCCCACACCGACTAC	363
sr-p70a	CCCCGACCCCCACCACTTTGAGGGTCACTTTCCAGGCCACGGCCAA	650
sr-p70f	CCCCGACCCCCACCACTTTGAGGGTCACTTTCCAGGCCACGGCCAA	372
sr-p70d	CCCCGACCCCCACCACTTTGAGGGTCACTTTCCAGGCCACGGCCAA	266
sr-p70e	CCCCGACCCCCACCACTTTGAGGGTCACTTTCCAGGCCACGGCCAA	266
sr-p70b	CCCCGACCCCCACCACTTTGAGGGTCACTTTCCAGGCCACGGCCAA	413
sr-p70a	GTCAGCCCCACCTGGACCGTACTCCCCGGCTCTTGAAAGAACCTCTACTGCCAGA	700
sr-p70f	GTCAGCCCCACCTGGACCGTACTCCCCGGCTCTTGAAAGAACCTCTACTGCCAGA	422
sr-p70d	GTCAGCCCCACCTGGACCGTACTCCCCGGCTCTTGAAAGAACCTCTACTGCCAGA	316
sr-p70e	GTCAGCCCCACCTGGACCGTACTCCCCGGCTCTTGAAAGAACCTCTACTGCCAGA	316
sr-p70b	GTCAGCCCCACCTGGACCGTACTCCCCGGCTCTTGAAAGAACCTCTACTGCCAGA	463
sr-p70a	TGGCCAAAGACATGCCACATCCAGATCAAGGTTCCACCCGGCCACCCCCA	750
sr-p70f	TGGCCAAAGACATGCCACATCCAGATCAAGGTTCCACCCGGCCACCCCCA	472
sr-p70d	TGGCCAAAGACATGCCACATCCAGATCAAGGTTCCACCCGGCCACCCCCA	366
sr-p70e	TGGCCAAAGACATGCCACATCCAGATCAAGGTTCCACCCGGCCACCCCCA	366
sr-p70b	TGGCCAAAGACATGCCACATCCAGATCAAGGTTCCACCCGGCCACCCCCA	513

FIG.15 cont.

24/36

sr-p70a	GGCCACCTGCCATCCGGCCATGCCCTGTTACAAAGAACCGGAGCGACCGTGAC	800
sr-p70f	GGCCACCTGCCATCCGGCCATGCCCTGTTACAAAGAACCGGAGCGACCGTGAC	522
sr-p70d	GGCCACCTGCCATCCGGCCATGCCCTGTTACAAAGAACCGGAGCGACCGTGAC	416
sr-p70e	GGCCACCTGCCATCCGGCCATGCCCTGTTACAAAGAACCGGAGCGACCGTGAC	416
sr-p70b	GGCCACCTGCCATCCGGCCATGCCCTGTTACAAAGAACCGGAGCGACCGTGAC	563
sr-p70a	CGAACCGTCCGTGAACCGCTGGCCCCAACCGAGCTCCGGGAGGGACTTCACCG	850
sr-p70f	CGAACCGTCCGTGAACCGCTGGCCCCAACCGAGCTCCGGGAGGGACTTCACCG	572
sr-p70d	CGAACCGTCCGTGAACCGCTGGCCCCAACCGAGCTCCGGGAGGGACTTCACCG	466
sr-p70e	CGAACCGTCCGTGAACCGCTGGCCCCAACCGAGCTCCGGGAGGGACTTCACCG	466
sr-p70b	CGAACCGTCCGTGAACCGCTGGCCCCAACCGAGCTCCGGGAGGGACTTCACCG	613
sr-p70a	AAGGACAGTCTGCTGCCAGCCCACCTCATCCGGGTGGGAAGGGCAATAAT	900
sr-p70f	AAGGACAGTCTGCTGCCAGCCCACCTCATCCGGGTGGGAAGGGCAATAAT	622
sr-p70d	AAGGACAGTCTGCTGCCAGCCCACCTCATCCGGGTGGGAAGGGCAATAAT	516
sr-p70e	AAGGACAGTCTGCTGCCAGCCCACCTCATCCGGGTGGGAAGGGCAATAAT	516
sr-p70b	AAGGACAGTCTGCTGCCAGCCCACCTCATCCGGGTGGGAAGGGCAATAAT	663
sr-p70a	CTCTCGGAGTGTGGATGCCCTGTCACCCGGCAGGAGCCGTCGGTGGCT	950
sr-p70f	CTCTCGGAGTGTGGATGCCCTGTCACCCGGCAGGAGCCGTCGGTGGCT	672
sr-p70d	CTCTCGGAGTGTGGATGCCCTGTCACCCGGCAGGAGCCGTCGGTGGCT	566
sr-p70e	CTCTCGGAGTGTGGATGCCCTGTCACCCGGCAGGAGCCGTCGGTGGCT	566
sr-p70b	CTCTCGGAGTGTGGATGCCCTGTCACCCGGCAGGAGCCGTCGGTGGCT	713
sr-p70a	GCCCCATTGAGCCACCCAGGGTGGGGACGGAAATTCACCAACCATCCCTGTA	1000
sr-p70f	GCCCCATTGAGCCACCCAGGGTGGGGACGGAAATTCACCAACCATCCCTGTA	722
sr-p70d	GCCCCATTGAGCCACCCAGGGTGGGGACGGAAATTCACCAACCATCCCTGTA	616
sr-p70e	GCCCCATTGAGCCACCCAGGGTGGGGACGGAAATTCACCAACCATCCCTGTA	616
sr-p70b	GCCCCATTGAGCCACCCAGGGTGGGGACGGAAATTCACCAACCATCCCTGTA	763

FIG. 15 cont.

25/36

sr-p70a	A C T T C A T G T G T A A C A G C A G G C T G T G T A G G G G G C A T G A A C C G G C G G C C C A T C	1050
sr-p70f	A C T T C A T G T G T A A C A G C A G G C T G T G T A G G G G G C A T G A A C C G G C G G C C C A T C	772
sr-p70d	A C T T C A T G T G T A A C A G C A G G C T G T G T A G G G G G C A T G A A C C G G C G G C C C A T C	666
sr-p70e	A C T T C A T G T G T A A C A G C A G G C T G T G T A G G G G G C A T G A A C C G G C G G C C C A T C	666
sr-p70b	A C T T C A T G T G T A A C A G C A G G C T G T G T A G G G G G C A T G A A C C G G C G G C C C A T C	813
sr-p70a	C T C A T C A T C A C C C T G G A G A T G C G G G A T T G G G C A G G T G C C T G G G C C G C C G	1100
sr-p70f	C T C A T C A T C A C C C T G G A G A T G C G G G A T T G G G C A G G T G C C T G G G C C G C C G	822
sr-p70d	C T C A T C A T C A C C C T G G A G A T G C G G G A T T G G G C A G G T G C C T G G G C C G C C G	716
sr-p70e	C T C A T C A T C A C C C T G G A G A T G C G G G A T T G G G C A G G T G C C T G G G C C G C C G	716
sr-p70b	C T C A T C A T C A C C C T G G A G A T G C G G G A T T G G G C A G G T G C C T G G G C C G C C G	863
sr-p70a	G T C C T T T G A G G G C C C A T C T G C G C C T G T C C T G G C C G A C C G A A A A G G C T G	1150
sr-p70f	G T C C T T T G A G G G C C C A T C T G C G C C T G T C C T G G C C G A C C G A A A A G G C T G	872
sr-p70d	G T C C T T T G A G G G C C C A T C T G C G C C T G T C C T G G C C G A C C G A A A A G G C T G	766
sr-p70e	G T C C T T T G A G G G C C C A T C T G C G C C T G T C C T G G C C G A C C G A A A A G G C T G	766
sr-p70b	G T C C T T T G A G G G C C C A T C T G C G C C T G T C C T G G C C G A C C G A A A A G G C T G	913
sr-p70a	A T G A G G A C C A C T A C C G G A G C A G G C A G G C A G G C T G A A C G A G G C T C C G C C A A G	1200
sr-p70f	A T G A G G A C C A C T A C C G G A G C A G G C A G G C A G G C T G A A C G A G G C T C C G C C A A G	922
sr-p70d	A T G A G G A C C A C T A C C G G A G C A G G C A G G C A G G C T G A A C G A G G C T C C G C C A A G	816
sr-p70e	A T G A G G A C C A C T A C C G G A G C A G G C A G G C A G G C T G A A C G A G G C T C C G C C A A G	816
sr-p70b	A T G A G G A C C A C T A C C G G A G C A G G C A G G C A G G C T G A A C G A G G C T C C G C C A A G	963
sr-p70a	A A C G G G G C C A G C C A A G C C A G C C C C C T G C C G C C G T C C G C C C C C	1250
sr-p70f	A A C G G G G C C A G C C A A G C C A G C C C C C T G C C G C C G T C C G C C C C C	972
sr-p70d	A A C G G G G C C A G C C A A G C C A G C C C C C T G C C G C C G T C C G C C C C C	866
sr-p70e	A A C G G G G C C A G C C A A G C C A G C C C C C T G C C G C C G T C C G C C C C C	866
sr-p70b	A A C G G G G C C A G C C A A G C C A G C C C C C T G C C G C C G T C C G C C C C C	1013

FIG. 15 cont.

26/36

sr-p70a	C G C C C T T G G G T C C G G T G T G A A G A A G C G G C A T G G A G A C G A C C G T	1300
sr-p70f	C G C C C T T G G G T C C G G T G T G A A G A A G C G G C A T G G A G A C G A C C G T	1022
sr-p70d	C G C C C T T G G G T C C G G T G T G A A G A A G C G G C A T G G A G A C G A C C G T	916
sr-p70e	C G C C C T T G G G T C C G G T G T G A A G A A G C G G C A T G G A G A C G A C C G T	916
sr-p70b	C G C C C T T G G G T C C G G T G T G A A G A A G C G G C A T G G A G A C G A C C G T	1063
sr-p70a	A C T A C C T T C A G G T G C C G A G G C C G G G A C T T T G A G A T C C T G A T G A A G C T G	1350
sr-p70f	A C T A C C T T C A G G T G C C G A G G C C G G G A C T T T G A G A T C C T G A T G A A G C T G	1072
sr-p70d	A C T A C C T T C A G G T G C C G A G G C C G G G A C T T T G A G A T C C T G A T G A A G C T G	966
sr-p70e	A C T A C C T T C A G G T G C C G A G G C C G G G A C T T T G A G A T C C T G A T G A A G C T G	966
sr-p70b	A C T A C C T T C A G G T G C C G A G G C C G G G A A C T T T G A G A T C C T G A T G A A G C T G	1113
sr-p70a	A A A G A G A G C C T G G A G C T G G A G T G G A G T G G C C A C T G G G T G G G A C T C	1400
sr-p70f	A A A G A G A G C C T G G A G C T G G A G T G G C C A C T G G G T G G G A C T C	1122
sr-p70d	A A A G A G A G C C T G G A G C T G G A G T G G C C A C T G G G T G G G A C T C	1016
sr-p70e	A A A G A G A G C C T G G A G C T G G A G T G G C C A C T G G G T G G G A C T C	1016
sr-p70b	A A A G A G A G C C T G G A G C T G G A G T G G C C A C T G G G T G G G A C T C	1163
sr-p70a	C T A T C G G G C A G C A G C C A G C C A G C C A G C C A G C C A G C C C T A C A G G C C C C	1450
sr-p70f	C T A T C G G G C A G C A G C C A G C C A G C C A G C C A G C C C T A C A G G C C C C	1172
sr-p70d	C T A T C G G G C A G C A G C C A G C C A G C C A G C C A G C C C T A C A G G C C C C	1066
sr-p70e	C T A T C G G G C A G C A G C C A G C C A G C C A G C C C T A C A G G C C - - - - -	1049
sr-p70b	C T A T C G G G C A G C A G C C A G C C A G C C A G C C C T A C A G G C C C T A C A G G C C C C	1213
sr-p70a	C G T C C T A C G G G C C G G T C C C T C T C G C C C A A G G T G G C A C G G G G G C A T G	1500
sr-p70f	C G T C C T A C G G G C C G G T C C C T C T C G C C C A A G G T G G C A C G G G G G C A T G	1222
sr-p70d	C G T C C T A C G G G C C G G T C C C T C T C G C C C A A G G T G G C A C G G G G G C A T G	1116
sr-p70e	- - - - -	1049
sr-p70b	C G T C C T A C G G G C C G G T C C C T C T C G C C C A A G G T G G C A C G G G G G C A T G	1263

FIG. 15 cont.

27/36

sr-p70a	A A C A A G C T G C C C T T C C G T C A A C C A G G C T G G G T G G G C C A G C A G	1550
sr-p70f	A A C A A G C T G C C C T T C C G T C A A C C A G G C T G G G T G G G C C A G C A G	1272
sr-p70d	A A C A A G C T G C C C T T C C G T C A A C C A G G C T G G G T G G G C C A G C A G	1166
sr-p70e	- - - - -	1049
sr-p70b	A A C A A G C T G C C C T T C C G T C A A C C A G G C T G G G T G G G C C A G C A G	1313
sr-p70a	T T C G G C A G G C T A C A C C C A A C C T G G G C C G T G G G C C C G G A T G C T C A A C A	1600
sr-p70f	T T C G G C A G G C T A C A C C C A A C C T G G G C C G T G G G C C C G G A T G C T C A A C A	1322
sr-p70d	T T C G G C A G G C T A C A C C C A A C C T G G G C C G T G G G C C C G G A T G C T C A A C A	1216
sr-p70e	- - - - -	1067
sr-p70b	T T C G G C A G G C T A C A C C C A A C C T G G G C C G T G G G C C C G G A T G C T C A A C A	1363
sr-p70a	A C C A T G G C C A C G G C A G T G C C A G C C A A C G G C A G G C A G G C A G C A G G C	1650
sr-p70f	A C C A T G G C C A C G G C A G T G C C A G C C A A C G G C A G G C A G G C A G G C A G G C	1372
sr-p70d	A C C A T G G C C A C G G C A G T G C C A G C C A A C G G C A G G C A G G C A G G C A G G C	1266
sr-p70e	A C C A T G G C C A C G G C A G T G C C A G C C A A C G G C A G G C A G G C A G G C A G G C	1117
sr-p70b	A C C A T G G C C A C G G C A G T G C C A G C C A A C G G C A G G C A G G C A G G C A G G C	1413
sr-p70a	G C C C A G T C C C A T G G T C T C G G G T C C C A C T G C A C T C C G C C A C C C C C T A C C A	1700
sr-p70f	G C C C A G T C C C A T G G T C T C G G G T C C C A C T G C A C T C C G C C A C C C C C T A C C A	1422
sr-p70d	G C C C A G T C C C A T G G T C T C G G G T C C C A C T G C A C T C C G C C A C C C C C T A C C A	1316
sr-p70e	G C C C A G T C C C A T G G T C T C G G G T C C C A C T G C A C T C C G C C A C C C C C T A C C A	1167
sr-p70b	G C C C A G T C C C A T G G T C T C G G G T C C C A C T G C A C T C C G C C A C C C C C T A C C A	1463
sr-p70a	C G C C G A C C C C A G C C C T C G T C A G T T T T A A C A G G A T T G G G T G T C C A A C C T	1750
sr-p70f	C G C C G A C C C C A G C C C T C G T C A G T T T T A A C A G G A T T G G G T G T C C A A C C T	1472
sr-p70d	C G C C G A C C C C A G C C C T C G T C A G T T T T A A C A G G A T T G G G T G T C C A A C C T	1366
sr-p70e	C G C C G A C C C C A G C C C T C G T C - - - - -	1186
sr-p70b	C G C C G A C C C C A G C C C T C G T C - - - - -	1482

FIG. 15 cont.

28/36

sr-p70a	GCATCGAGTATTCA	CCCTCCC	AAGGTTACAGAACCTG	CAG	1800
sr-p70f	GCATCGAGTATTCA	CCCTCCC	AAGGTTACAGAACCTG	CAG	1522
sr-p70d	GCATCGAGTATTCA	CCCTCCC	AAGGTTACAGAACCTG	CAG	1416
sr-p70e	-	-	-	-	1186
sr-p70b	-	-	-	-	1482
sr-p70a	AACCTGACCAATTGAGGACCT	TGGGG	CCCTGAAAGATCCCCGAGCAGTACCCG	1850	
sr-p70f	AACCTGACCAATTGAGGACCT	TGGGG	CCCTGAAAGATCCCCGAGCAGTACCCG	1572	
sr-p70d	AACCTGACCAATTGAGGACCT	TGGGG	CCCTGAAAGATCCCCGAGCAGTACCCG	1466	
sr-p70e	-	-	-	-	1223
sr-p70b	-	-	-	-	1519
sr-p70a	CATGACCCATCTGGCC	GGG	GGGCTGGAAAGGAGGAA	GACTACA	1900
sr-p70f	CATGACCCATCTGGCC	GGG	GGGCTGGAAAGGAGGAA	GACTACA	1622
sr-p70d	CATGACCCATCTGGCC	GGG	GGGCTGGAAAGGAGGAA	GACTACA	1516
sr-p70e	CATGACCCATCTGGCC	GGG	GGGCTGGAAAGGAGGAA	GACTACA	1273
sr-p70b	CATGACCCATCTGGCC	GGG	GGGCTGGAAAGGAGGAA	GACTACA	1569
sr-p70a	GCACCGGGAGCA	GGCTGGCT	GGCAACGGCTAGCTTCA	CCATTC	1950
sr-p70f	GCACCGGGAGCA	GGCTGGCT	GGCAACGGCTAGCTTCA	CCATTC	1672
sr-p70d	GCACCGGGAGCA	GGCTGGCT	GGCAACGGCTAGCTTCA	CCATTC	1566
sr-p70e	GCACCGGGAGCA	GGCTGGCT	GGCAACGGCTAGCTTCA	CCATTC	1323
sr-p70b	GCACCGGGAGCA	GGCTGGCT	GGCAACGGCTAGCTTCA	CCATTC	1619
sr-p70a	GGCGGGCTCA	GGGAAC	GGGCACTGGCATGGGAGGCC	ACTTT	2000
sr-p70f	GGCGGGCTCA	GGGAAC	GGGCACTGGCATGGGAGGCC	ACTTT	1722
sr-p70d	GGCGGGCTCA	GGGAAC	GGGCACTGGCATGGGAGGCC	ACTTT	1616
sr-p70e	GGCGGGCTCA	GGGAAC	GGGCACTGGCATGGGAGGCC	ACTTT	1373
sr-p70b	GGCGGGCTCA	GGGAAC	GGGCACTGGCATGGGAGGCC	ACTTT	1669

FIG. 15 cont.

29/36

sr-p70a	CCGGCGTGCACCAATCACCATCCCCAACCCGCGGCCAGGGCGCC	2050
sr-p70f	CCGGCGTGCACCAATCACCATCCCCAACCCGCGGCCAGGGCGCC	1772
sr-p70d	CCGGCGTGCACCAATCACCATCCCCAACCCGCGGCCAGGGCGCC	1666
sr-p70e	CCGGCGTGCACCAATCACCATCCCCAACCCGCGGCCAGGGCGCC	1423
sr-p70b	CCGGCGTGCACCAATCACCATCCCCAACCCGCGGCCAGGGCGCC	1719
sr-p70a	GCCCCTTGACGAGTGGGGACATTCCGGACCTGGCCTTCGAACTGGCAAGGCC	2100
sr-p70f	GCCCCTTGACGAGTGGGGACATTCCGGACCTGGCCTTCGAACTGGCAAGGCC	1822
sr-p70d	GCCCCTTGACGAGTGGGGACATTCCGGACCTGGCCTTCGAACTGGCAAGGCC	1716
sr-p70e	GCCCCTTGACGAGTGGGGACATTCCGGACCTGGCCTTCGAACTGGCAAGGCC	1473
sr-p70b	GCCCCTTGACGAGTGGGGACATTCCGGACCTGGCCTTCGAACTGGCAAGGCC	1769
sr-p70a	CGGCAAGCCAAGCCATCAAGGGAGGTTCACGGAGGGCGAGATCCACCTGA	2150
sr-p70f	CGGCAAGCCAAGCCATCAAGGGAGGTTCACGGAGGGCGAGATCCACCTGA	-
sr-p70d	CGGCAAGCCAAGCCATCAAGGGAGGTTCACGGAGGGCGAGATCCACCTGA	-
sr-p70e	CGGCAAGCCAAGCCATCAAGGGAGGTTCACGGAGGGCGAGATCCACCTGA	-
sr-p70b	CGGCAAGCCAAGCCATCAAGGGAGGTTCACGGAGGGCGAGATCCACCTGA	-
sr-p70a	GCCTCGCCCTGGCTGGCAGGCCCTGGCCACCGGCCAGAGGACCCAGGCC	2200
sr-p70f	- - - - -	-
sr-p70d	- - - - -	-
sr-p70e	- - - - -	-
sr-p70b	- - - - -	-
sr-p70a	CCCTCTCCCTTGTGTCAAAACCTGGCTCAGGAGGGCAGGACCTTCGG	2250
sr-p70f	- - - - -	-
sr-p70d	- - - - -	-
sr-p70e	- - - - -	-
sr-p70b	- - - - -	-

FIG.15 cont.

30/36

sr-p70a	G C T G T G C C C G G G A A G G C A A G G T C C C G C C A T C C C C A G G C A C C T C A C A G	2300
sr-p70f	- - - - -	- - - - -
sr-p70d	- - - - -	- - - - -
sr-p70e	- - - - -	- - - - -
sr-p70b	- - - - -	- - - - -
		1817

sr-p70a	G C C C C A G G A A A G G G C C A G G C C A G C C G A A G C C G C C T G T G G A C A G C C T G A G T C A	2350
sr-p70f	- - - - -	- - - - -
sr-p70d	- - - - -	- - - - -
sr-p70e	- - - - -	- - - - -
sr-p70b	- - - - -	- - - - -
		1817

sr-p70a	C C T G C A G A A C C	2361
sr-p70f	- - - - -	1870
sr-p70d	- - - - -	1764
sr-p70e	- - - - -	1521
sr-p70b	- - - - -	1817

FIG. 15 cont.

31/36

sr-p70a-	MAQSTATSPDGCTTFEHLWSSLEPDSTYFDLPOSSRGNNNEVVGCTDSSMD	50
sr-p70f-	-	2
sr-p70d-	-	2
sr-p70b-	MAQSTATSPDGCTTFEHLWSSLEPDSTYFDLPOSSRGNNNEVVGCTDSSMD	50
sr-p70e-	-	1
sr-p70a-	V F H L E G M T T S V M A Q F N L L S S T M D Q M S S R A A S A S P Y T P E H A A S V P T H S P Y A	100
sr-p70f-	V F H L E G M T T S V M A Q F N L L S S T M D Q M S S R A A S A S P Y T P E H A A S V P T H S P Y A	52
sr-p70d-	L Y V G D P A R H L A T A Q F N L L S S T M D Q M S S R A A S A S P Y T P E H A A S V P T H S P Y A	51
sr-p70b-	V F H L E G M T T S V M A Q F N L L S S T M D Q M S S R A A S A S P Y T P E H A A S V P T H S P Y A	100
sr-p70e-	L Y V G D P A R H L A T A Q F N L L S S T M D Q M S S R A A S A S P Y T P E H A A S V P T H S P Y A	51
sr-p70a-	Q P S S T F D T M S P A P V I P S N T D Y P G P H H F E V T F Q Q S S T A K S A T W T Y S P L L K K	150
sr-p70f-	Q P S S T F D T M S P A P V I P S N T D Y P G P H H F E V T F Q Q S S T A K S A T W T Y S P L L K K	102
sr-p70d-	Q P S S T F D T M S P A P V I P S N T D Y P G P H H F E V T F Q Q S S T A K S A T W T Y S P L L K K	101
sr-p70b-	Q P S S T F D T M S P A P V I P S N T D Y P G P H H F E V T F Q Q S S T A K S A T W T Y S P L L K K	150
sr-p70e-	Q P S S T F D T M S P A P V I P S N T D Y P G P H H F E V T F Q Q S S T A K S A T W T Y S P L L K K	101
sr-p70a-	L Y C Q I A K T C P I Q I K V S T P P P P G T A I R A M P V Y K K A E H V T D V V K R C P N H E L G	200
sr-p70f-	L Y C Q I A K T C P I Q I K V S T P P P P G T A I R A M P V Y K K A E H V T D V V K R C P N H E L G	152
sr-p70d-	L Y C Q I A K T C P I Q I K V S T P P P P G T A I R A M P V Y K K A E H V T D V V K R C P N H E L G	151
sr-p70b-	L Y C Q I A K T C P I Q I K V S T P P P P G T A I R A M P V Y K K A E H V T D V V K R C P N H E L G	200
sr-p70e-	L Y C Q I A K T C P I Q I K V S T P P P P G T A I R A M P V Y K K A E H V T D V V K R C P N H E L G	151
sr-p70a-	R D F N E G Q S A P A S H L I R V E G N N L S Q Y V D D P V T G R Q S V V P Y E P P Q V G T E F T	250
sr-p70f-	R D F N E G Q S A P A S H L I R V E G N N L S Q Y V D D P V T G R Q S V V P Y E P P Q V G T E F T	202
sr-p70d-	R D F N E G Q S A P A S H L I R V E G N N L S Q Y V D D P V T G R Q S V V P Y E P P Q V G T E F T	201
sr-p70b-	R D F N E G Q S A P A S H L I R V E G N N L S Q Y V D D P V T G R Q S V V P Y E P P Q V G T E F T	250
sr-p70e-	R D F N E G Q S A P A S H L I R V E G N N L S Q Y V D D P V T G R Q S V V P Y E P P Q V G T E F T	201

FIG. 16

32 / 36

sr-p70a_ T I L Y N F M C N S S C V G G M N R R P I L I I I T L E M R D G Q V L G R R S S F E G R I C A C P G R 300
 sr-p70f_ T I L Y N F M C N S S C V G G M N R R P I L I I I T L E M R D G Q V L G R R S S F E G R I C A C P G R 252
 sr-p70d_ T I L Y N F M C N S S C V G G M N R R P I L I I I T L E M R D G Q V L G R R S S F E G R I C A C P G R 251
 sr-p70b_ T I L Y N F M C N S S C V G G M N R R P I L I I I T L E M R D G Q V L G R R S S F E G R I C A C P G R 300
 sr-p70e_ T I L Y N F M C N S S C V G G M N R R P I L I I I T L E M R D G Q V L G R R S S F E G R I C A C P G R 251

sr-p70a_ DR K A D E D H Y R E Q Q A L N E S S A K N G A A S K R A F K Q S P P A V P A L G A G V K K R R H G 350
 sr-p70f_ DR K A D E D H Y R E Q Q A L N E S S A K N G A A S K R A F K Q S P P A V P A L G A G V K K R R H G 302
 sr-p70d_ DR K A D E D H Y R E Q Q A L N E S S A K N G A A S K R A F K Q S P P A V P A L G A G V K K R R H G 301
 sr-p70b_ DR K A D E D H Y R E Q Q A L N E S S A K N G A A S K R A F K Q S P P A V P A L G A G V K K R R H G 350
 sr-p70e_ DR K A D E D H Y R E Q Q A L N E S S A K N G A A S K R A F K Q S P P A V P A L G A G V K K R R H G 301

sr-p70a_ D E D T Y Y L Q V R G R E N F E I L M K L K E S S L E L M E L V P Q P L V D S Y R Q Q Q Q L L Q R P S 400
 sr-p70f_ D E D T Y Y L Q V R G R E N F E I L M K L K E S S L E L M E L V P Q P L V D S Y R Q Q Q Q L L Q R P S 352
 sr-p70d_ D E D T Y Y L Q V R G R E N F E I L M K L K E S S L E L M E L V P Q P L V D S Y R Q Q Q Q L L Q R P S 351
 sr-p70b_ D E D T Y Y L Q V R G R E N F E I L M K L K E S S L E L M E L V P Q P L V D S Y R Q Q Q Q L L Q R P S 400
 sr-p70e_ D E D T Y Y L Q V R G R E N F E I L M K L K E S S L E L M E L V P Q P L V D S Y R Q Q Q Q L L Q R P S 351

sr-p70a_ H L Q P P S Y G P V L S P M N K V H G G M N K L P S V N Q L V G Q P P P H S S A A T P N L G P V G P 450
 sr-p70f_ H L Q P P S Y G P V L S P M N K V H G G M N K L P S V N Q L V G Q P P P H S S A A T P N L G P V G P 402
 sr-p70d_ H L Q P P S Y G P V L S P M N K V H G G M N K L P S V N Q L V G Q P P P H S S A A T P N L G P V G P 401
 sr-p70b_ H L Q P P S Y G P V L S P M N K V H G G M N K L P S V N Q L V G Q P P P H S S A A T P N L G P V G P 450
 sr-p70e_ R D A Q Q P W P - - - - - R S A S Q R R D E Q Q P Q R P V - - - - - 375

sr-p70a_ G M L N N H G H A V P A N G E M S S S H S A Q S M V S G S H C T P P P Y H A D P S S L V S F L T G L 500
 sr-p70f_ G M L N N H G H A V P A N G E M S S S H S A Q S M V S G S H C T P P P Y H A D P S S L V S F L T G L 452
 sr-p70d_ G M L N N H G H A V P A N G E M S S S H S A Q S M V S G S H C T P P P Y H A D P S S L V S F L T G L 451
 sr-p70b_ G M L N N H G H A V P A N G E M S S S H S A Q S M V S G S H C T P P P Y H A D P S S L V S F L T G L 499
 sr-p70e_ - - - - - H G L G V P L - - - - - H S A T P L L P R P Q P R - - - - - 395

FIG. 16 cont.

33/36

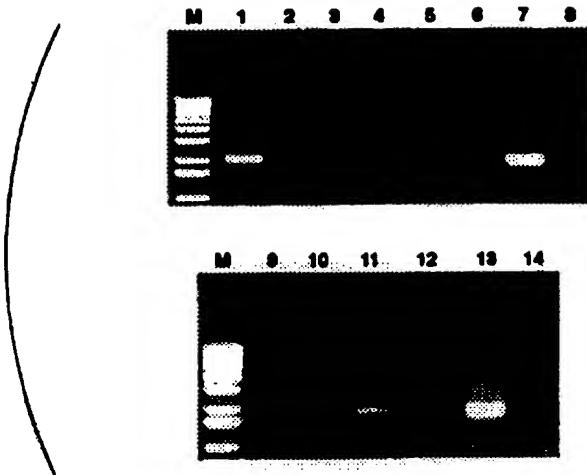
sr-p70a-	GC P N C I E Y F T S Q G L Q S S I Y H L O N L T I E D L G A L K I P E Q Y R M T I W R G L Q D L K Q	550
sr-p70f-	GC P N C I E Y F T S Q G L Q S S I Y H L O N L T I E D L G A L K I P E Q Y R M T I W R G L Q D L K Q	502
sr-p70d-	GC P N C I E Y F T S Q G L Q S S I Y H L O N L T I E D L G A L K I P E Q Y R M T I W R G L Q D L K Q	501
sr-p70b-	-	-
sr-p70e-	-	-
	Q D L G A L K I P E Q Y R M T I W R G L Q D L K Q	499
	Q D L G A L K I P E Q Y R M T I W R G L Q D L K Q	420
sr-p70a-	G H D Y S T A Q Q L L R S S N A A T I S I G G S G E L Q R V M E A V H F R V R H T I T I P N R G	600
sr-p70f-	G H D Y S T A Q Q L L R S S N A A T I S I G G S G E L Q R V M E A V H F R V R H T I T I P N R G	552
sr-p70d-	G H D Y S T A Q Q L L R S S N A A T I S I G G S G E L Q R V M E A V H F R V R H T I T I P N R G	551
sr-p70b-	-	-
sr-p70e-	G H D Y S T A Q Q L L R S S N A A T I S I G G S G E L Q R V M E A V H F R V R H T I T I P N R G	499
	Q D L G A L K I P E Q Y R M T I W R G L Q D L K Q	470
sr-p70a-	G P G G G P D E W A D F G F D L P D C K A R K Q P I K E E F T E A E I H	636
sr-p70f-	G P G G G P D E W A D F G F D L P D C K A R K Q P I K E E F T E A E I H	588
sr-p70d-	G P G G G P D E W A D F G F D L P D C K A R K Q P I K E E F T E A E I H	587
sr-p70b-	-	-
sr-p70e-	G P G G G P D E W A D F G F D L P D C K A R K Q P I K E E F T E A E I H	499
	Q D L G A L K I P E Q Y R M T I W R G L Q D L K Q	506

FIG.16 cont.

1 TAACGCCCGGGGCCCCCTACTCCCCGGGGCCTCCCTCCCCGCCCATATAACCCGC 60
 61 CTAGGGGGCGGGGCAAGCCCCCTGCTTCCCAGGGCGACCCGGCCGGAGGCTCGCG 120
 121 CCCGCGAAGGGGACCGAGGAAACCGGGCCCGGCCAGGGCAGGGCACGGACGCCGA 180
 181 TGCCCCGGGGCTGGGACGGCTGAGCTGCCAGGGAGCTGGAGGGCGGGAAAGATG 240
 M 1
 241 GCCCAGTCCACCGCCACCTCCCTGATGGGGGACCCACGTTTGAGCACCTCTGGAGCTCT 300
 2 A Q S T A T S P D G G T T F E H L W S S 21
 301 CTGGAACAGAACAGACCTACTTCGACCCAGTCAGTCAGCCGGGGAAATAATGAGGTG 360
 22 L E P D S T Y F D L P Q S S R G N N E V 41
 361 GTGGGGCGGAACGGATTCAGCATGGACGTCTTCACCTGGAGGGCATGactACNtCTGTC 420
 42 V G G T D S S M D V F H L E G M T T S V 61
 421 ATGGCCCAGTTCAATCTGCTGAGCAGGACCATATGGACCATGAGCAATGAGCG 480
 62 M A Q F N L L S S T M D Q M S S R A A S 81
 481 GCCAGCCCCCTACACCCAGAGCACGGCGCCAGGGCTGCCACCCACTCGCCCTACGCACAA 540
 82 A S P Y T P E H A A S V P T H S P Y A Q 101
 541 CCCAGCTCCACCTTCGACACCATTGTCATCCCCCTCCAAACACGGACTAC 600
 102 P S S T F D T M S P A P V I P S N T D Y 121
 601 CCCGGACCCACCACTTTGAGGTCACTTTCAGCAGTCCAGCAGGGCAAGTCAGGCCACC 660
 122 P G P H H F E V T F Q Q S S T A K S A T 141
 661 TGGACGTA.....
 142 W T

FIG. 17

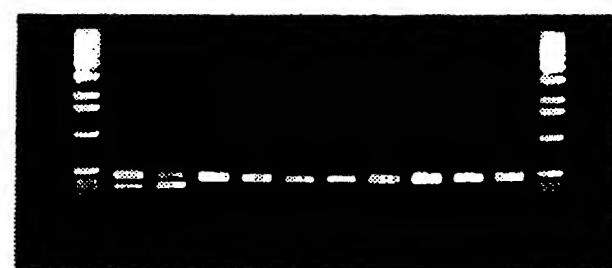
35 / 36

FIG. 18

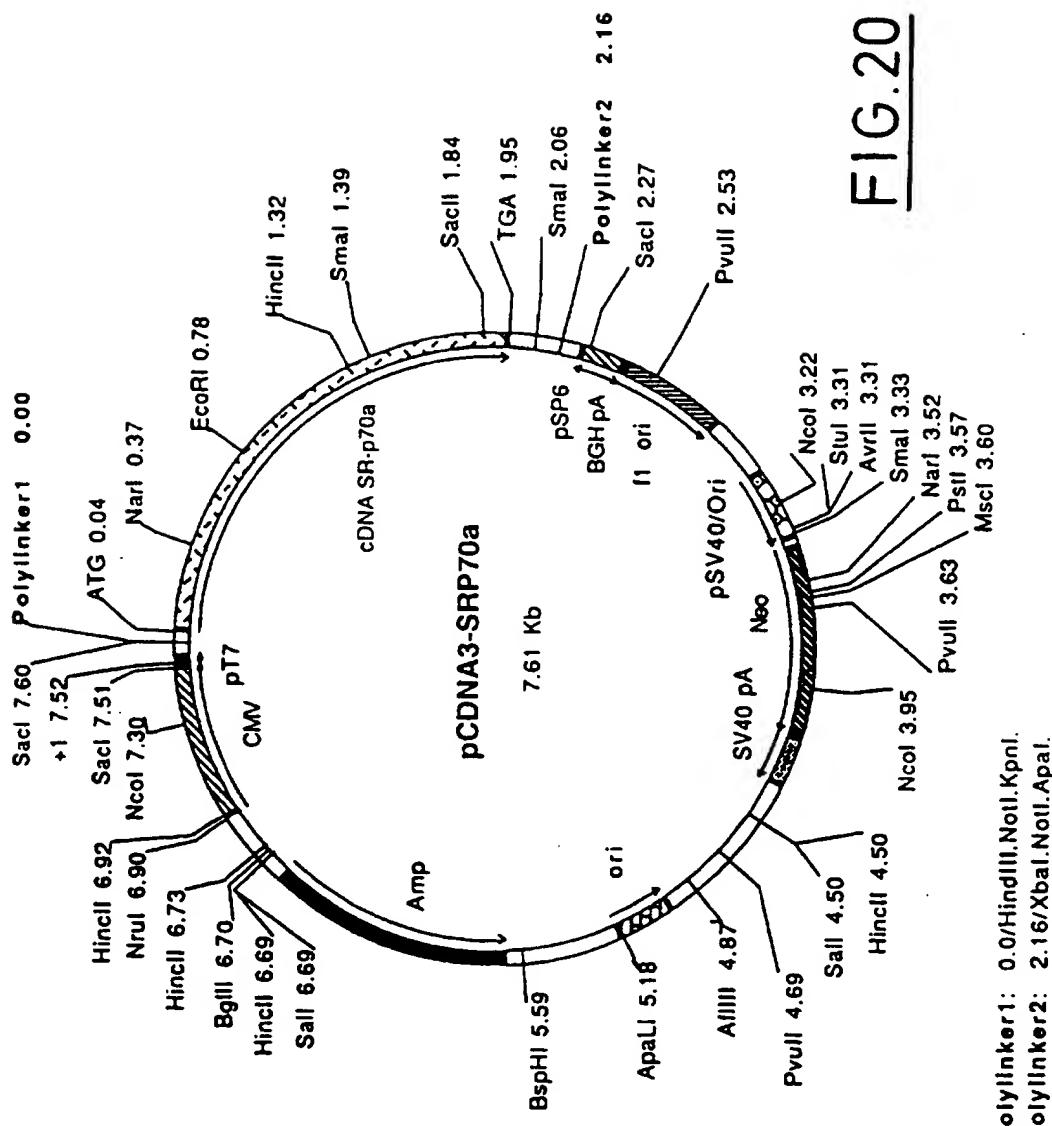
M 1 2 3 4 5 6 7 8 9 10 M

FIG. 19A

M 1 2 3 4 5 6 7 8 9 10 M

FIG. 19 B

FEUILLE DE REMPLACEMENT (REGLE 26)



INTERNATIONAL SEARCH REPORT

International Application No

PCT/FR 97/00214

A. CLASSIFICATION OF SUBJECT MATTER				
IPC 6	C07K14/47	C12N15/12	C12Q1/68	A61K39/395
				G01N33/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C07K A61K C12N C12Q G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	- / --	

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

* Special categories of cited documents :

- *'A' document defining the general state of the art which is not considered to be of particular relevance
- *'E' earlier document but published on or after the international filing date
- *'L' document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *'O' document referring to an oral disclosure, use, exhibition or other means
- *'P' document published prior to the international filing date but later than the priority date claimed

- *'T' later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *'X' document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *'Y' document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- *'&' document member of the same patent family

Date of the actual completion of the international search

12 June 1997

Date of mailing of the international search report

20.06.97

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax (+31-70) 340-3016

Authorized officer

Gac, G

INTERNATIONAL SEARCH REPORT

In...national Application No
PCT/FR 97/00214

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SCIENCE, vol. 237, 1987, pages 1620-1624, XP000604718 BODRUG: "Molecular analysis of a constitutional X-autosome translocation in a female with muscular dystrophy" see page 1622; figure 4	13,14
X	& DATABASE STRAND ref. EMHUM hsrtmdl, AN: L08092 6 April 1993 see sequence	13,14, 16,17
A X	& J. MOL. BIOL., vol. 232, no. 1, 1993, pages 314-321, XP000604618 MC NAUGHTON ET AL.: "A cluster of transposon-like repetitive sequences in intron 7 of the human dystrophin gene" see the whole document	22,23 13,14
A	---	18,22,23
X	NUCLEIC ACID RESEARCH, vol. 16, no. 23, 1988, page 1183 XP002014924 SOUSSI ET AL.: "Nucleotide sequence of a cDNA encoding the chicken p53 nuclear protein" see the whole document	13,14
A X	& DATABASE STRAND ref. Swissprot: P53-chick : AN : P10360 1 March 1989 see sequence	1-9, 13-17,25 13,14
A	---	1-9, 13-17
X	WO 94 01563 A (ENERGY BIOSYSTEMS CORPORATION) 20 January 1994 see sequence nA 1	13,14
A	& DATABASE STRAND ref. Pat-SA93-D: SA77122 see sequence nA 1	16,17, 22,23
	---	-/-

INTERNATIONAL SEARCH REPORT

	International Application No PCT/FR 97/00214
--	---

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	BIOCHEM. BIOPHYS. RES. COMMUN., vol. 194, no. 2, 30 July 1993, pages 698-705, XP002014925 IWASE ET AL.: "Identification of protein-tyrosine kinase genes preferentially expressed in embryo stomach and gastric cancer" see page 700; figure 1 & DATABASE STRAND ref. EMHUM1: Hserklip, AN: D37827 16 August 1994 see sequence : ---	13,14
X	EP 0 377 295 A (ELI LILLY AND COMPANY) 11 July 1990 see page 6 & DATABASE STRAND ref. Tpsd-D: I08282, AN: I08282 18 February 1995 see sequence ---	13,14
X	DATABASE STRAND ref. EN960713, AN: Z75711 XP002014926 see the sequence & NATURE, vol. 368, 1994, pages 32-38, WILSON ET AL.: "2.2. Mb of contiguous nucleotide sequence from chromosome III of C. elegans" ---	13,14
X	FR 2 692 594 A (PEREZ J-C.) 24 December 1993 & DATABASE GENESEQ AN=Q55626, 12 July 1994 XP002014927 see the alignment of the sequences ---	13,14
X	WO 94 08241 A (DEUTSCHE KREBSFORCHUNGSZENTRUM STIFTUNG DES ÖFFENTLICHES RECHTS) 14 April 1994 see the whole document	13,14
A	---	1-12, 15-36
	-/-	

INTERNATIONAL SEARCH REPORT

International Application No
PCT/FR 97/00214

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	MOL. CELL. BIOL., vol. 6, no. 9, September 1986, pages 3232-3239, XP000604639 ARAI ET AL.: "Immunologically distinct p53 molecules generated by alternative splicing" see the whole document	13,14
A	& DATABASE STRAND ref. Pir2: S38822 13 January 1995 see sequence ---	3,4
A	NATURE GENETICS, vol. 6, no. 4, April 1994, pages 357-362, XP000604628 NEUMANN ET AL.: "Multifactorial inheritance of neural tube defects : localization of the major gene and recognition of modifiers in ct mutant genes" see the whole document	13,17, 22,23
X	DATABASE EMBL ID: CEF26F12, AC= U55373, XP002032930 see the alignment of the sequences	13,14
A	& NATURE, vol. 368, 1994, pages 32-38, WILSON ET AL.: "2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans"	16
X	DATABASE EMBL ID=AC= S77819, 29 September 1995 XP002032931 see the alignment of the sequences & CANCER LETTERS, vol. 92, no. 2, 8 June 1995, pages 181-186, XP000674693 KRAEGEL ET AL.: "Sequence analysis of canine p53 in the region of exons 3-8"	13,14
	---	-/-

INTERNATIONAL SEARCH REPORT

International Application No PCT/FR 97/00214

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>DATABASE EMBL ID: SIP53, AC=M75145, 24 August 1991 XP002032932 see the alignment of the sequences & GENE, vol. 112, 1992, pages 241-245, DE FROMENTEL ET AL.: "Rainbow trout p53: cDNA cloning and biochemical characterization" see the whole document ---</p>	13,14
X	<p>DATABASE EMBL ID: BTP53, AC= X81704, 14 December 1994 XP002032933 see the alignment of the sequences & DNA SEQ., vol. 5, no. 4, 1995, pages 261-264, XP000674685 DESQUIEDT ET AL.: "Nucleotide sequence of bovine p53 tumor-suppressor cDNA" see the whole document ---</p>	13,14
A	<p>NUCLEIC ACIDS RES., vol. 20, no. 8, 1992, pages 1879-1882, XP002032929 GRYAZNOV ET AL.: "Selective O-phosphitilation with nucleoside phosphoramidite reagents" see page 1880 ---</p>	18
A	<p>INT. J. RADIAT. BIOL. RELAT. STUD. PHYS., CHEM. MED., vol. 51, no. 3, 1987, pages 429-439, XP000674638 TEOULE ET AL.: "Gamma-irradiation of homodeoxyoligonucleotides 32P-labelled at one end : computer simulation of the chain length distribution of the radioactive fragments" see page 426 -----</p>	18

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/FR 97/00214

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO 9401563 A	20-01-94	AU 4671893 A		31-01-94
		CA 2139876 A		20-01-94
		CN 1085254 A		13-04-94
		EP 0651808 A		10-05-95
		JP 7507691 T		31-08-95
		NO 950081 A		09-03-95
		US 5356801 A		18-10-94
		US 5578478 A		26-11-96
<hr/>				
EP 377295 A	11-07-90	AT 118042 T		15-02-95
		AU 622253 B		02-04-92
		AU 4709889 A		28-06-90
		CA 2005649 A		22-06-90
		DE 68920987 D		16-03-95
		DE 68920987 T		22-06-95
		ES 2067556 T		01-04-95
		HU 208713 B		28-12-93
		JP 2227082 A		10-09-90
<hr/>				
FR 2692594 A	24-12-93	NONE		
<hr/>				
WO 9408241 A	14-04-94	EP 0614531 A		14-09-94
		JP 7501711 T		23-02-95
<hr/>				

RAPPORT DE RECHERCHE INTERNATIONALE

D : de Internationale No
PCT/FR 97/00214

A. CLASSEMENT DE L'OBJET DE LA DEMANDE
CIB 6 C07K14/47 C12N15/12 C12Q1/68 A61K39/395 G01N33/68

Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB

B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE

Documentation minimale consultée (système de classification suivi des symboles de classement)
CIB 6 C07K A61K C12N C12Q G01N

Documentation consultée autre que la documentation minimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche

Base de données électronique consultée au cours de la recherche internationale (nom de la base de données, et si cela est réalisable, termes de recherche utilisés)

C. DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
	- / --	

Voir la suite du cadre C pour la fin de la liste des documents

Les documents de familles de brevets sont indiqués en annexe

* Catégories spéciales de documents cités:

- *A* document définissant l'état général de la technique, non considéré comme particulièrement pertinent
- *E* document antérieur, mais publié à la date de dépôt international ou après cette date
- *L* document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)
- *O* document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens
- *P* document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée

- *T* document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention
- *X* document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré seulement
- *Y* document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier
- *&* document qui fait partie de la même famille de brevets

3

Date à laquelle la recherche internationale a été effectivement achevée 12 Juin 1997	Date d'expédition du présent rapport de recherche internationale 30.06.97
Nom et adresse postale de l'administration chargée de la recherche internationale Office Européen des Brevets, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+ 31-70) 340-3016	Fonctionnaire autorisé Gac, G

RAPPORT DE RECHERCHE INTERNATIONALE

L'Ande Internationale No
PCT/FR 97/00214

C(suite) DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications vistes
X	SCIENCE, vol. 237, 1987, pages 1620-1624, XP000604718 BODRUG: "Molecular analysis of a constitutional X-autosome translocation in a female with muscular dystrophy" voir page 1622; figure 4	13,14
X	& DATABASE STRAND ref. EMHUM hsrtmdl, AN: L08092 6 Avril 1993 voir séquence	13,14, 16,17
A		22,23
X	& J. MOL. BIOL., vol. 232, no. 1, 1993, pages 314-321, XP000604618 MC NAUGHTON ET AL.: "A cluster of transposon-like repetitive sequences in intron 7 of the human dystrophin gene" voir le document en entier	13,14
A	---	18,22,23
X	NUCLEIC ACID RESEARCH, vol. 16, no. 23, 1988, page 1183 XP002014924 SOUSSI ET AL.: "Nucleotide sequence of a cDNA encoding the chicken p53 nuclear protein" voir le document en entier	13,14
A		1-9, 13-17,25
X	& DATABASE STRAND ref. Swissprot: P53-chick : AN : P10360 1 Mars 1989 voir séquence	13,14
A	---	1-9, 13-17
X	WO 94 01563 A (ENERGY BIOSYSTEMS CORPORATION) 20 Janvier 1994 voir séquence nA 1	13,14
A		16,17, 22,23
	& DATABASE STRAND ref. Pat-SA93-D: SA77122 voir séquence nA 1 ---	
	-/-	

RAPPORT DE RECHERCHE INTERNATIONALE

D. *nde Internationale No
PCT/FR 97/00214*

C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS		
Catégorie *	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	BIOCHEM. BIOPHYS. RES. COMMUN., vol. 194, no. 2, 30 Juillet 1993, pages 698-705, XP002014925 IWASE ET AL.: "Identification of protein-tyrosine kinase genes preferentially expressed in embryo stomach and gastric cancer" voir page 700; figure 1 & DATABASE STRAND ref. EMHUM1: Hserk1p, AN: D37827 16 Août 1994 voir séquence ---	13,14
X	EP 0 377 295 A (ELI LILLY AND COMPANY) 11 Juillet 1990 voir page 6 & DATABASE STRAND ref. Tpsd-D: I08282, AN: I08282 18 Février 1995 voir séquence ---	13,14
X	DATABASE STRAND ref. EN960713, AN: Z75711 XP002014926 voir séquence & NATURE, vol. 368, 1994, pages 32-38, WILSON ET AL.: "2.2. Mb of contiguous nucleotide sequence from chromosome III of C. elegans" ---	13,14
X	FR 2 692 594 A (PEREZ J-C.) 24 Décembre 1993 & DATABASE GENESEQ AN=Q55626, 12 Juillet 1994 XP002014927 voir alignement des séquences ---	13,14
X	WO 94 08241 A (DEUTSCHES KREBSFORSCHUNGZENTRUM STIFTUNG DES ÖFFENTLICHES RECHTS) 14 Avril 1994 voir le document en entier	13,14
A	---	1-12, 15-36
	-/-	

RAPPORT DE RECHERCHE INTERNATIONALE

L. Ande Internationale No PCT/FR 97/00214
--

C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS		no. des revendications visées
Catégorie	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	
X	MOL. CELL. BIOL., vol. 6, no. 9, Septembre 1986, pages 3232-3239, XP000604639 ARAI ET AL.: "Immunologically distinct p53 molecules generated by alternative splicing" voir le document en entier	13,14
A	& DATABASE STRAND ref. Pir2: S38822 13 Janvier 1995 voir séquence ---	3,4
A	NATURE GENETICS, vol. 6, no. 4, Avril 1994, pages 357-362, XP000604628 NEUMANN ET AL.: "Multifactorial inheritance of neural tube defects : localization of the major gene and recognition of modifiers in ct mutant genes" voir le document en entier ---	13,17, 22,23
X	DATABASE EMBL ID: CEF26F12, AC= U55373, XP002032930 voir alignements des séquences	13,14
A	& NATURE, vol. 368, 1994, pages 32-38, WILSON ET AL.: "2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans" ---	16
X	DATABASE EMBL ID=AC= S77819, 29 Septembre 1995 XP002032931 voir alignements des séquences & CANCER LETTERS, vol. 92, no. 2, 8 Juin 1995, pages 181-186, XP000674693 KRAEGEL ET AL.: "Sequence analysis of canine p53 in the region of exons 3-8" ---	13,14
		-/-

RAPPORT DE RECHERCHE INTERNATIONALE

Demande Internationale No PCT/FR 97/00214
--

C.(suite) DOCUMENTS CONSIDÉRÉS COMME PERTINENTS

Catégorie	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	<p>DATABASE EMBL ID: SIP53, AC=M75145, 24 Août 1991 XP002032932 voir alignements des séquences & GENE, vol. 112, 1992, pages 241-245, DE FROMENTEL ET AL.: "Rainbow trout p53: cDNA cloning and biochemical characterization" voir le document en entier ---</p>	13,14
X	<p>DATABASE EMBL ID: BTP53, AC= X81704, 14 Décembre 1994 XP002032933 voir alignements des séquences & DNA SEQ., vol. 5, no. 4, 1995, pages 261-264, XP000674685 DESQUIEDT ET AL.: "Nucleotide sequence of bovine p53 tumor-suppressor cDNA" voir le document en entier ---</p>	13,14
A	<p>NUCLEIC ACIDS RES., vol. 20, no. 8, 1992, pages 1879-1882, XP002032929 GRYAZNOV ET AL.: "Selective O-phosphitilation with nucleoside phosphoramidite reagents" voir page 1880 ---</p>	18
A	<p>INT. J. RADIAT. BIOL. RELAT. STUD. PHYS., CHEM. MED., vol. 51, no. 3, 1987, pages 429-439, XP000674638 TEOULE ET AL.: "Gamma-irradiation of homodeoxyoligonucleotides 32P-labelled at one end : computer simulation of the chain length distribution of the radioactive fragments" voir page 426 -----</p>	18

RAPPORT DE RECHERCHE INTERNATIONALE

Renseignements relatifs aux membres de familles de brevets

Demande Internationale No
PCT/FR 97/00214

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
WO 9401563 A	20-01-94	AU 4671893 A CA 2139876 A CN 1085254 A EP 0651808 A JP 7507691 T NO 950081 A US 5356801 A US 5578478 A	31-01-94 20-01-94 13-04-94 10-05-95 31-08-95 09-03-95 18-10-94 26-11-96
EP 377295 A	11-07-90	AT 118042 T AU 622253 B AU 4709889 A CA 2005649 A DE 68920987 D DE 68920987 T ES 2067556 T HU 208713 B JP 2227082 A	15-02-95 02-04-92 28-06-90 22-06-90 16-03-95 22-06-95 01-04-95 28-12-93 10-09-90
FR 2692594 A	24-12-93	AUCUN	
WO 9408241 A	14-04-94	EP 0614531 A JP 7501711 T	14-09-94 23-02-95